

UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II
DIPARTIMENTO INGEGNERIA DEI MATERIALI E DELLA PRODUZIONE



DOTTORATO DI RICERCA IN INGEGNERIA
DEI MATERIALI E DELLA PRODUZIONE
indirizzo in
Tecnologie e Sistemi di Produzione

XX ciclo
TESI DI DOTTORATO

**TECNICHE DI OTTIMIZZAZIONE
DELLA CAPACITA' PRODUTTIVA
NELLE AZIENDE DI SERVIZI**

COORDINATORE
PROF. CH.MO
ING. GIUSEPPE GIORLEO

DOTTORANDO
ING. ANNARITA FRANCO

TUTOR
CH.MO. PROF. ING. LIBERATINA C. SANTILLO

ANNO ACCADEMICO 2006-2007

CAP 1: LE AZIENDE DI SERVIZI: VINCOLI E CRITICITA'

1.1	Definizione e caratteristiche principali dei servizi	4
1.2	Tipologie di servizi	5
1.3	Il ruolo del cliente	7
1.3.1	Il valore percepito	8
1.3.2	Prezzi e Value driver	10
1.3.3	I rischi della non differenziazione	11
1.3.4	I vincoli della differenziazione	13
1.3.5	Il triangolo del servizio	16
1.3.6	La qualità nei servizi	17
1.3.6.1	Misurazione della produttività nei servizi	18
1.3.7	La dimensione tempo	20
1.4	Classificazione dei servizi in funzione del grado di interazione col cliente	21
1.5	Servizi front-office e back-office	22
1.5.1	Disaccoppiamento o integrazione front e back office	24
1.6	L'operations management nei servizi	27
1.6.1	Sviluppo storico dell'OM	28
1.6.2	Operations management e tendenze attuali	29

CAP 2: IL BILANCIO DELLA CAPACITA' PRODUTTIVA IN BACK OFFICE: LA SCHEDULAZIONE DEL PERSONALE

2.1.	Labor scheduling	32
2.2.	Prima fase: determinazione della richiesta del servizio	33
2.2.1.	Passo 1: determinare la natura del lavoro	34
2.2.2.	Passo 2: identificare i fattori che influenzano il lavoro	35
2.2.3.	Passo 3: determinare se i fattori chiave del lavoro variano nel breve periodo	35
2.2.4.	Passo 4: determinare l'intervallo di tempo opportuno per tracciare la variazione dei labor drivers	37
2.2.5.	Passo 5: prevedere la variazione temporale dei labor drivers	38
2.2.6.	Passo 6: ridurre la variazione casuale con tecniche di smoothing	40
2.2.7.	Passo 7: controllare, attraverso misurazioni e rilevamenti, l'accuratezza della previsione	42
2.2.8.	Passo 8: definire una finestra di esecuzione per un lavoro controllabile	43
2.3.	Seconda fase: traduzione della richiesta del servizio nella domanda di impiegati	43
2.3.1.	Confronto tra i tre approcci usati per la traduzione	45
2.3.2.	Sviluppo di uno standard economico	51
2.3.3.	Aspetti fondamentali per l'applicazione di uno standard economico	57
2.4.	Terza fase: lo scheduling degli impiegati	60
2.4.1.	I sistemi classici di "workforce scheduling"	60
2.4.2.	I nuovi modelli	63
2.4.3.	Sviluppo dello scheduling della forza lavoro	64
2.4.4.	Risoluzione dei problemi di scheduling	66
2.5.	Quarta fase: adattamento dello scheduling effettuato alla reale richiesta del servizio	70
2.5.1.	Un approccio per prevedere la richiesta del servizio in un dato giorno	71
2.5.2.	Azioni da sviluppare in tempo reale e importanza del "cross-training"	81

CAP 3: IL BILANCIO DELLA CAPACITA' PRODUTTIVA IN FRONT OFFICE: LA GESTIONE DELLE CODE

3.1 Centri di contatto e fenomeni di congestione	84
3.2 I passi fondamentali nella teoria delle code	86
3.3 Sistema di servizio	87
3.3.1 Strutturazione di un sistema di servizio	88
3.3.2 Notazione di kendall	92
3.3.3 misure di prestazione e analisi stazionaria	93
3.3.4 Teorema di little	97
3.3.5 Proprietà pasta	98
3.4 Processi stocastici di arrivo e di servizio	100
3.4.1 Il ruolo della distribuzione esponenziale	100
3.4.2 Il processo di poisson	103
3.5 Caratterizzazione dei processi di nascita e morte nei sistemi di servizio	107
3.6 Misure di prestazione dei principali sistemi di code	113
3.6.1 Sistemi a singolo servente m/m/1	115
3.6.2 Sistemi multiserventi m/m/s	120
3.7 introduzione al modello palm/erlang-a.	123
3.7.1 La pazienza del cliente in un sistema di code	129
3.8 Code visibili e code invisibili: l'importanza dell'informazione	132
3.8.1 "Modello di reneging" e "modello di balking"	134

CAP 4: PROTECTION LEVEL E TECNICHE DI OVERBOOKING. CASO APPLICATIVO

4.1 Introduzione	137
4.2 Requisiti necessari per l'applicazione dello yield management	138
4.3 Applicazioni	140
4.4 Strumenti dello yield management	141
4.4.1 Segmentazione del mercato e pricing	141
4.4.2 Previsione della domanda	144
4.4.3 Allocazione della capacità	149
4.4.4 Prenotazione e vendita	152
4.5 Analisi dello Yield management	154
4.5.1 Vantaggi	154
4.5.2 Pregiudizi e critiche	157
4.6. La gestione delle risorse	159
4.7 Concetto di Protection Level	160
4.7.1 Metodi risolutivi nel caso di due tariffe	161
4.7.2 Analisi deterministica	162
4.7.3 Analisi tramite albero delle decisioni	163
4.7.4 Analisi marginale	166
4.7.5 Metodo risolutivo nel caso di più di due tariffe	167

4.8 Concetto di overbooking	172
4.9 Disciplina normativa	179
4.10 Metodi risolutivi	182
4.10.1 Analisi sui valori medi	182
4.10.2 Analisi tramite foglio di calcolo	182
4.10.3 Analisi tramite albero delle decisioni	185
4.10.4 Analisi marginale	186
4.10.5 Modello del ricavo atteso	187
4.10.6 Modello di Karaesmen e van Ryzin	188
4.11 Esempio applicativo: il caso di un ente di formazione	194
4.11.1 L'ente di formazione Ares e l'offerta formativa	194
4.11.2 L'implementazione del metodo	197
4.12 Gestione dell'overbooking attraverso un software implementato in Microsoft Access	204
4.12.1 Come funziona il software	211
4.13 Conclusioni	224
<i>Indice Tabelle</i>	225
<i>Indice Figure</i>	226
<i>Bibliografia</i>	227

CAPITOLO I

NATURA E CLASSIFICAZIONE DEI SERVIZI

In questo capitolo, saranno esposte, innanzitutto, le principali caratteristiche dei servizi, mostrando analogie e differenze dal settore manifatturiero. La trattazione si sposterà poi su quella che in effetti costituisce la caratteristica peculiare di un servizio: il contatto con il cliente. Si focalizzerà quindi l'attenzione sul ruolo che egli occupa all'interno del sistema-servizio e sull'integrazione della sua esperienza nel processo di sviluppo del servizio offerto. Dopo aver esaminato il fattore tempo, come parametro che incide in maniera considerevole sulla performance e sulla qualità del sistema produttivo in questione e alle tecniche di natura psicologica utilizzate per fronteggiare i suoi effetti negativi, si passa alla classificazione dei servizi in front e back office, le cui caratteristiche sono funzione proprio del grado d'interazione suddetto. Il capitolo si chiude quindi con una panoramica sull'Operations Management, ovvero l'insieme di tutte quelle attività aziendali volte alla progettazione, realizzazione e miglioramento di tutti i sistemi d'impresa che producono beni o erogano servizi.

1.1 Definizione e caratteristiche principali dei servizi

Ogni sistema di produzione impiega risorse per convertire fattori in entrata (input) in fattori in uscita (output) desiderati. Questi processi di trasformazione si realizzano in campi diversi, e ciò che li differenzia l'uno dall'altro è la natura del proprio output. Il *servizio* rappresenta “l'output *intangibile* del processo produttivo”, le cui caratteristiche quindi non si identificano con la fisicità e materialità caratterizzanti l'output della produzione industriale (bene tangibile).

Il servizio, quindi, può essere definito come uno scambio di attività fra un soggetto che esprime bisogni, attese, desideri e un altro soggetto che fornisce risposte. In tutti questi casi, quindi, il cliente potrebbe percepire la sensazione di non essere entrato in possesso di “qualcosa di materiale”, non riuscendo così a cogliere il valore di ciò che ha pagato: è indispensabile pertanto, dare valore al servizio, considerandolo strumento di soluzione dei problemi e di risposta a esigenze specifiche espresse da parte del consumatore.

Un altro aspetto da sottolineare, è quello legato alla *impossibilità di stoccaggio*, caratteristica che rende il processo produttivo difficile da standardizzare e tempificare. In questo contesto infatti, il prodotto/servizio deve essere consumato all'istante, non può essere conservato per un uso futuro; il processo e il prodotto infatti, vengono sviluppati simultaneamente (il processo è il prodotto), il che comporta di conseguenza la necessità che la capacità produttiva debba essere immediatamente disponibile per produrre un servizio quando occorre e localizzata preferibilmente nel punto in cui si trova il cliente al momento della richiesta del servizio stesso. Nel settore manifatturiero infatti, la distribuzione delle merci al cliente segue il processo di produzione del bene fisico; nel caso dei servizi invece accade l'opposto: inizialmente deve essere distribuita al cliente la capacità di erogare il servizio (fisicamente o attraverso un mezzo di comunicazione come il telefono), quindi può essere prodotto il servizio.

Un'altra peculiarità di un sistema che eroga servizi, è rappresentata da un alto grado di *volatilità della domanda*, significativamente superiore rispetto ai sistemi di produzione manifatturiera. Ciò accade innanzitutto per l'impossibilità di stoccaggio suddetta, che non permette di accumulare scorte nei periodi di stanca per poi utilizzarle nei periodi di picco della domanda, il che comporta una certa difficoltà nel mantenere un livello di produzione più o meno stabile; in secondo luogo questo stato di variabilità è dettato anche dalla natura "personalizzata" del servizio, il cui processo di erogazione ha come destinatari clienti diversi che hanno spesso esigenze diverse, che nel corso del processo, sperimentano diversi livelli di esperienza e possono richiedere diversi quantitativi di transazioni, comportando in questo modo una certa variabilità nei tempi di erogazione del servizio stesso. A causa di questa volatilità, spesso i servizi pianificano l'aumento di capacità per periodi ridotti, compresi fra i 10 e 30 minuti, contro gli incrementi di periodo settimanale, tipici della produzione industriale [1].

È indispensabile inoltre evidenziare il fatto che i servizi non sono necessariamente soggetti tutti alle medesime regole (quel che funziona bene nell'erogazione di un certo tipo di servizio può rivelarsi disastroso per un altro), che alla qualità del lavoro eseguito non corrisponde necessariamente la qualità del servizio, e che spesso per una gestione efficace, risultano molto utili conoscenze nell'ambito della gestione della produzione, del marketing e della gestione del personale; spesso infatti, la possibilità di buona riuscita del servizio, è funzione del grado di addestramento acquisito dal personale a volte ancor prima di entrare a far parte dell'azienda.

Pertanto "l'industria del servizio" deve essere costituita di tante attività volte a manipolare oggetti fisici, soggetti umani e informazioni che in qualche modo si influenzino l'un l'altro, senza che subiscano però una vera e propria trasformazione fisica; l'attenzione deve essere rivolta quindi all'uso dei prodotti fisici e alla loro funzionalità relativamente alle suddette attività, piuttosto che al processo di trasformazione fisica che li coinvolge.

1.2 Tipologie di servizi

Il luogo di produzione e il grado di coinvolgimento diretto del cliente nella creazione dell'output, sovente fondamentali per i servizi e non per i beni, come descritto precedentemente, consentono di effettuare una suddivisione delle aziende erogatrici di servizi in [1]:

- 1. Imprese di servizi:** si tratta di organizzazioni la cui attività primaria richiede un certo grado di interazione con la clientela per produrre il servizio. All'interno di questa categoria possiamo operare un'ulteriore, fondamentale, distinzione:

- **Servizi sul campo (field-based services):** sono per definizione quelli dotati di un sistema di distribuzione fisica che giunge fino al consumatore;
- **Servizi “in situ” (facilities-based services):** in questo caso è il cliente che deve recarsi in “fabbrica” per potere usufruire del servizio.

In molti contesti la tecnologia ha consentito di passare dalla erogazione di servizi in sede a quella sul campo.

2. Servizi interni: si tratta della gestione di tutti quei servizi necessari a supportare le attività di grandi organizzazioni; ne fanno parte l'*Information Technology*, contabilità, manutenzione.

Risulta difficile comunque effettuare una vera e propria distinzione tra il settore dei servizi e quello manifatturiero, in quanto si verifica sempre più una sorta di integrazione dei due, ben evidenziata dal fatto che alcune aziende manifatturiere offrono spesso numerosi **servizi integrati nel prodotto**; questi si possono distinguere in:

- 1. Servizi core:** sono rappresentati da quei prodotti costruiti correttamente, personalizzati in base alle necessità del cliente, consegnati con puntualità e a prezzi competitivi, rispettando i classici obiettivi di prestazione perseguiti dalla funzione produzione (qualità, flessibilità, velocità e prezzo).
- 2. Servizi a valore aggiunto:** sono quei servizi volti a “facilitare la vita del consumatore”. Nell’ambito della fabbrica, questo tipo di servizi possono essere classificati in quattro ampie categorie:
 - **Informazioni:** si riferiscono alla capacità di fornire dati significativi su prestazioni del prodotto, parametri di processo e costi sia a soggetti interni, per esempio al reparto Ricerca e Sviluppo per poter migliorare le sue modalità operative, sia a clienti esterni per poter migliorare l’utilizzo dei prodotti. Per esempio, il reparto Qualità di Hewlett-Packard a Fort Collins [1], distribuisce ai rappresentanti e ai tecnici di assistenza, statistiche aggiornate e videocassette che riportano i test di qualità svolti e le effettive prestazioni dei prodotti offerti, rilevate sul campo.
 - **Problem Solving:** rappresenta la capacità di aiutare i soggetti interni ed esterni a risolvere i problemi, specie sul versante qualità.
 - **Supporto alle vendite:** consiste nella capacità di potenziare l’azione del settore commerciale e del marketing, attraverso dimostrazioni delle attrezzature, della tecnologia e dei sistemi di produzione che l’azienda intende vendere.
 - **Assistenza sul campo:** si configura come capacità di sostituire rapidamente le parti difettose o di reintegrare celermente le scorte per evitare ritardi o *stockout*.

I servizi a valore aggiunto erogati ai clienti esterni generano due vantaggi: innanzitutto differenziano l'impresa dai suoi concorrenti e in secondo luogo permettono la realizzazione di rapporti positivi tra clientela e azienda.

Esistono inoltre imprese di servizi che producono esse stesse i prodotti fisici che consegnano al cliente (McDonald's da questo punto di vista rappresenta un classico esempio di come, pur producendo un bene tangibile, sia possibile far ricadere la sua attività nella categoria dei servizi, poiché il completamento del processo di produzione avviene attraverso una forma di contatto col cliente[1]).

Da quanto appena detto circa l'integrazione che negli ultimi anni si è venuta a creare tra il settore manifatturiero e quello dei servizi, risulta ovviamente sbagliato pensare ad un sistema di servizio solo come un sistema ad alta intensità di manodopera (labor intensive) e a quello manifatturiero solo ad alta intensità di capitale e tecnologia (capital intensive); a tal proposito si riporta il caso significativo della "First National City Bank"[4], una delle più grandi banche a livello mondiale che annovera circa 37.000 impiegati tra le sue fila, di cui più della metà a contatto diretto col pubblico, ma con una grande percentuale che lavora "dietro le quinte" e costituisce la "fabbrica" dell'azienda; così come l'IBM, la più grande azienda manifatturiera a livello mondiale che però ha con sé circa 270.000 dipendenti, di cui oltre la metà avente a che fare con la clientela, sia nelle vendite che nell'assistenza. È sempre più evidente quindi che tante aziende manifatturiere tecnologicamente avanzate, senza una efficace attività di servizi di supporto alla produzione, non raggiungerebbero mai il livello di vendite fissato.

1.3 Il ruolo del cliente

Il periodo successivo alle due grandi guerre è stato caratterizzato, in termini economici, dal prevalere della domanda sull'offerta[12]. Non era necessario quindi preoccuparsi più di tanto del cliente: l'importante era avere prodotti da vendere.

Negli anni '70-'80, con l'avvento delle prime crisi energetiche, che fecero sì che fossero introdotti i programmi di "austerità", si invertì la tendenza, e le aziende iniziarono ad affinare i loro sistemi di attrazione dei clienti. Si ebbero, in questo periodo, i primi programmi strutturati di marketing e le società di consulenza teorizzarono e proposero le 4 P: prodotto/servizio, promozione, prezzo, presenza.

Con gli anni '90, la competizione si è inasprita, sono caduti i cartelli e i protezionismi, la tecnologia è diventata largamente disponibile, si sono sviluppati nuovi e più differenziati stili di vita. Le aziende hanno reagito con interventi di marketing più mirati; il mercato non è più qualcosa di indistinto, ma è costituito da diversi "segmenti" omogenei per atteggiamenti, comportamenti, stili di acquisto.

Il XXI secolo è caratterizzato dalla globalizzazione: nasce quindi la necessità di guardare al singolo cliente per personalizzare, o meglio, per dare la possibilità alle aziende di personalizzare i prodotti/servizi, la promozione, il prezzo, la presenza. La personalizzazione del servizio diventa il fattore di differenziazione più efficace e quindi il cliente tende sempre più a rappresentare il punto focale di tutte le decisioni e azioni intraprese dall'azienda.

1.3.1 Il valore percepito

Nella ricerca dell'affermazione competitiva, l'impresa deve fornire ai clienti motivi di preferenza della propria offerta rispetto a quella dei concorrenti, motivi che possono essere connessi alla maggiore valore del rapporto prezzo/ prestazioni: quanto più ampie e consolidate sono le ragioni della preferenza, tanto più forte è la posizione dell'impresa sul mercato. La superiorità competitiva dell'offerta deve essere riferita al concetto di valore percepito dal cliente in quanto quest'ultimo, in sede di acquisto, sviluppa un processo valutativo il cui punto di arrivo è l'acquisizione di un sistema di attributi a fronte del sostenimento di un determinato onere/sacrificio.

La differenziazione deve quindi essere attuata, nell'ambito del confronto con i concorrenti, ricercando la soddisfazione dei clienti e l'equilibrio nella relazione con il prezzo e i costi aziendali [17].

E' opportuno sottolineare che mentre la differenziazione ha generalmente un impatto sui costi dell'impresa, in quanto richiede un consumo di risorse specifico che si traduce nella generazione di costi aziendali, esso ha un effetto più incerto e articolato sul fronte dei ricavi. Incerto in quanto correlato alla rilevanza per il cliente e alla percezione da parte dello stesso delle azioni sviluppate; articolato in quanto può essere di breve termine o di lungo termine e può tradursi nell'incremento:

- Di prezzo, e quindi nell'aumento dei ricavi unitari
- Delle frequenze di consumo, e quindi dei volumi relativamente ai clienti attuali
- Del numero di clienti, ovvero nell'acquisizione di nuovi clienti
- Delle motivazioni alla base della fedeltà da parte dei clienti, cioè nella stabilità dei ricavi nel lungo periodo

La definizione delle politiche di differenziazione nell'ottica del valore per il cliente è il risultato di un processo complesso e articolato che richiede di :

- Individuare i value driver, ovvero le componenti dell'offerta più rilevanti (determinanti del valore) per i clienti;
- Effettuare una segmentazione della domanda in ottica value based

- Analizzare la superiorità/inferiorità dell'offerta in raffronto a quella dei concorrenti
- Individuare le opportunità di differenziazione in relazione all'esistenza dei diversi segmenti di clienti
- Verificare l'esistenza di vincoli che possono costituire un limite alle scelte di differenziazione dell'impresa

I primi quattro punti assumono validità nella generalità delle situazioni orientate allo sviluppo di attività di differenziazione; l'ultimo presenta rilevanti elementi di peculiarità con riferimento ai servizi [18].

1.3.2 Prezzi e value driver

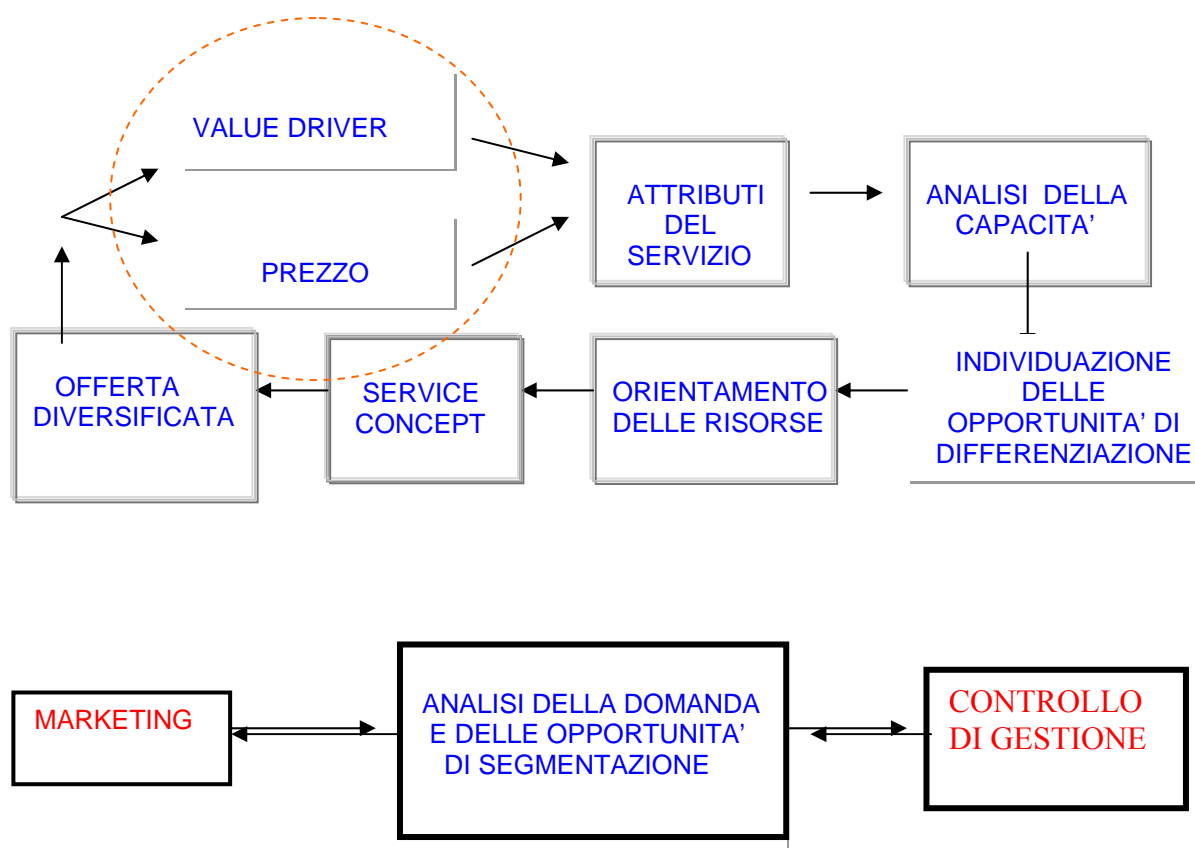


Figura 1.1 Punto di partenza per l'integrazione

L'analisi dell'atteggiamento della domanda e delle opportunità di segmentazione della stessa consiste nell'individuazione delle componenti dell'offerta, e, tra queste, dei value driver, ovvero delle determinanti del valore-utilità per il cliente [17].

Quest'ultimo è subordinato alla ricostruzione della customer value chain [19], ossia dei nessi cognitivi attraverso i quali il cliente collega gli attributi che compongono i beni e i servizi offerti dall'impresa ai benefici che gli stessi possono offrire. Il primo passo consiste nell'identificazione dei benefici ricercati dal cliente e degli attributi /componenti dell'offerta ad essi connessi. Cioè per ognuna delle categorie di componenti che descrivono il servizio possono essere individuati gli attributi all'origine delle prestazioni. La mappatura dei componenti/attributi dell'offerta richiede di ripercorrere l'intera esperienza sviluppata dal cliente nella fruizione di un determinato servizio, dalla fase di pre-acquisto alla conclusione del

rapporto con l'impresa. Accanto agli attributi relativi al servizio, devono essere considerati anche altri elementi che entrano nel rapporto con il mercato come ad esempio la comunicazione impersonale e l'immagine, che possono essere relative all'azienda e allo specifico servizio.

Il riferimento al prezzo completa l'analisi: il prezzo ha una funzione duplice, di indicatore della qualità del servizio e di controprestazione monetaria, a fronte dell'acquisizione dell'utilità connessa all'insieme degli attributi del servizio.

Dopo l'individuazione dei value driver, occorre analizzare l'importanza degli attributi per il cliente: si tratta di individuare gli attributi che, alla luce della loro connessione con i benefici ricercati dal cliente e con i relativi costi di acquisizione, determinano il valore che il cliente può trarre dal bene/servizio. Questa rilevazione consente di sviluppare una segmentazione value-based, aggregando i clienti che presentano le più significative similarità in termini di sistema di preferenza nei confronti di un determinato servizio. L'analisi deve essere completata con riguardo al livello di soddisfazione espresso dai clienti nei confronti dei diversi attributi, raffrontato al giudizio espresso sulle prestazioni offerte dai concorrenti.

L'identificazione dei value driver e la conoscenza degli atteggiamenti dei clienti nei confronti degli attributi dell'offerta dell'impresa, in raffronto a quella dei concorrenti, consentono di individuare le opportunità di differenziare e di orientare le risorse aziendali. La verifica della coerenza tra sistemi di preferenze/percezioni e impieghi di risorse dell'impresa evidenzia le opportunità di configurazione e di eventuale modifica delle attività -potenziamento, riduzione, riconversione e così via- e di conseguenza, della struttura dei costi aziendali.

In particolare, l'analisi dell'importanza degli attributi / componenti è l'elemento centrale per la verifica della coerenza tra la struttura dell'offerta e benefici ricercati dai clienti: rappresenta il riferimento per l'analisi della coerenza tra benefici ricercati e impieghi di risorse aziendali. La conoscenza della valutazione delle prestazioni, orientata a rilevare la percezione del cliente relativamente alle prestazioni ricevute, anche in raffronto alle alternative proposte dai concorrenti, ha la funzione di evidenziare le modalità con cui determinate prestazioni sono erogate, ossia l'efficacia dell'impiego delle risorse con riferimento allo specifico attributo.

1.3.3 I rischi della non differenziazione

La differenziazione dell'offerta è lo strumento fondamentale per acquisire le preferenze dei clienti. Politiche volte a mantenere un'uniformità di offerta, in presenza di eterogeneità delle determinanti del valore espresse dalla domanda, possono generare alcuni rischi che nel caso dei servizi, sono riconducibili ai seguenti fattori:

- Connotazione poco chiara del posizionamento del servizio e scarsa capacità di attrazione dei clienti target
- ‘Sovvenzionamento’ incrociato tra segmenti di clientela
- Non corretto dimensionamento della capacità produttiva

La scarsa attenzione alle differenze connesse all’eterogeneità dei benefici ricercati dai clienti può determinare la definizione di un profilo di offerta medio: in questo caso, i rischi sono determinati da un posizionamento poco chiaro, che tende ad aggregare i clienti con esigenze differenziate. In presenza di una concorrenza elevata, il rischio di una connotazione poco chiara del posizionamento può tradursi in uno scarso interesse da parte dei clienti.

A riguardo, poi, delle scelte di posizionamento, è necessario evidenziare il fatto che il tema della differenziazione deve essere affrontato a due livelli:

- come risposta ad una scelta di fondo riferita all’analisi di segmentazione del mercato
- come successiva ricerca di un adattamento delle caratteristiche dell’offerta in risposta a parametri di microsegmentazione.

Al primo livello, si tratta di individuare una selezione delle caratteristiche di base dell’offerta necessarie per operare in una determinata area di business. Su questa base avviene poi una scelta di segmentazione a livello micro, che orienta le decisioni di affinamento principale dell’offerta e di individuazione di quelle secondarie.

Proprio con riferimento a questo secondo livello, la definizione delle caratteristiche dell’offerta può anche avvenire secondo un approccio di adattamento progressivo e in via sperimentale, sino al punto in cui la tipologia e i volumi di clienti attratti siano all’origine di un margine considerato ottimale dal management aziendale. Si tratta di una realtà che può essere riscontrabile con riferimento a diverse tipologie di servizi, e segnatamente a quelle che si traducono in un’offerta costituita da un mix di attributi destinati alla soddisfazione di un sistema di benefici articolati. In questi casi infatti può rivelarsi opportuno adottare un approccio dinamico, di verifica e di ridefinizione continua dei sistemi di attributi dell’offerta, a fronte di una continua analisi dei benefici ricercati e del monitoraggio del comportamento dei clienti.

Passando a considerare il secondo rischio derivante dal mancato riconoscimento di diversi sistemi di preferenze dei singoli attributi, ossia il fenomeno del <<sovvenzionamento incrociato>> tra clienti/segmenti, è opportuno sottolineare che la proposta di un’offerta media, rivolta ad un cliente medio è in realtà espressione di una riduzione semplificatrice della realtà effettuata dall’impresa. La scelta può reggere sul mercato fino a quando un’impresa concorrente non si inserisce in un ambito più specifico, con un rapporto prestazioni/prezzo rivolto ad una particolare categoria di clienti e fondato sulla ricerca di una

corrispondenza maggiore tra utilizzo effettivo delle prestazioni offerte ed il prezzo relativo. In questo caso è probabile che il nuovo entrante si rivolga ai clienti penalizzati dal rapporto descritto, ossia a clienti che fruiscono solo in parte delle prestazioni offerte e che, in ragione dell'uniformità di prezzo, di fatto <<sovvenzionano>> coloro che utilizzano il servizio con maggiore intensità.

Infine, la mancanza di logiche di personalizzazione può essere all'origine di una definizione non ottimale della capacità produttiva, e quindi del rischio di determinare sprechi e livelli di sottoutilizzo delle strutture e delle risorse aziendali, da un lato; colli di bottiglia e disservizi per eccessi di domanda, dall'altro.

1.3.4 Vincoli alla differenziazione

Il perseguimento di politiche di adattamento dei servizi alle diverse tipologie di clienti deve avvenire nel rispetto di alcuni vincoli, connessi all'erogazione, che possono essere di due tipi:

- 1. Vincoli che configurano situazioni di impossibilità di adattamento del rapporto prestazioni/prezzo a differenti clienti, in ragione della contestualità di fruizione.**
- 2. Vincoli connessi alla disponibilità di risorse (personale, competenze, strutture)**

1) Vincoli connessi alla fruizione contestuale

Nell'ambito dei servizi emergono situazioni caratterizzate da gradi di libertà differenti in termini di configurazione dei livelli di personalizzazione del rapporto prestazioni/prezzo nei confronti dei diversi clienti.

I vincoli che condizionano le scelte di differenziazione sono connessi alla personalizzazione del servizio, alle strutture fisiche e al tempo.

- **Vincoli di personalizzazione del singolo servizio** Alcuni servizi richiedono una determinata configurazione di risorse o di competenze per l'erogazione di prestazioni che sono necessariamente a favore della generalità dei clienti. Si tratta di solito di servizi necessari, o anche secondari, ma connessi da uno stretto rapporto di funzionalità con quello principale, la cui configurazione diviene necessariamente comune nei confronti di tutti i clienti. In quest'ottica è rilevante la distinzione tra:

- 1.** servizi che possono essere adattati con riferimento a specifici gruppi di clienti nell'ambito di quelli obiettivo;
- 2.** servizi che sono definibili in modo univoco per la generalità dei clienti ai quali è rivolta l'offerta.

Con riferimento alla prima tipologia di servizi, la differenziazione può essere utilizzata come opportunità per definire un'offerta rispondente ad esigenze che possono essere specifiche solo di una parte di clienti serviti, al fine di ottenere una disponibilità degli stessi a sostenere una variazione di prezzo incrementale.

Un esempio può essere rappresentato dal servizio di trasporto e consegna a domicilio di prodotti acquistati in un punto di vendita.

Nel caso invece di servizi la cui offerta è necessariamente destinata alla generalità di clienti, questa situazione non si può verificare. A titolo di esempio, si pensi al caso in cui un gruppo di clienti dimostri di apprezzare un aggiornamento delle informazioni contenute in una banca dati in tempi più rapidi rispetto a quelli attuali e di essere disposto a pagare un prezzo più elevato per ottenere tale prestazione. La scelta di adattare l'offerta al cluster individuato si traduce necessariamente in un vantaggio per la generalità dei clienti e pone quindi un problema in termini di scelte di prezzo più elevato che non tutti i clienti sono disposti a pagare, né d'altra parte la relazione tra prestazioni e prezzo può essere gestita in termini di focalizzazione sul singolo segmento/cliente, poiché non si potrebbero giustificare differenziali di prezzo sulla base di una fruizione di prestazioni omogenee.

- **Vincoli connessi alle strutture fisiche:** Si tratta di vincoli connessi alla dimensione spaziale delle strutture fisiche adibite all'erogazione del servizio, che richiamano le problematiche tipiche dei vincoli di capacità produttiva, aggravate dalla non immagazzinabilità dei servizi. Tali vincoli in alcuni casi sono connessi al limite dato dallo spazio disponibile, in altri al fatto che tali strutture prevedono modalità di fruizione comuni ed omogenee da parte dei clienti.

Le strutture fisiche nell'ambito delle quali si sviluppa l'erogazione presentano il limite della dimensione spaziale. Si tratta del vincolo più evidente che delimita la possibilità di realizzare offerte alternative o servizi dedicati a specifici clienti. Si pensi al caso di un villaggio turistico, che potrebbe avere l'opportunità di arricchire la gamma di servizi offerti ma che trova proprio nello spazio disponibile il limite che costringe ad una selezione tra gli stessi.

- **Vincoli connessi al tempo:** Si tratta di vincoli legati alla domanda (ad esempio ai momenti di tempo libero in cui è possibile usufruire dei servizi di intrattenimento) o all'offerta (si pensi all'impossibilità di andare oltre un certo periodo di apertura dei punti vendita o di strutture adibite all'erogazione di servizi al pubblico). A tal proposito deve essere sottolineato il fatto che il tempo è un vincolo del tutto peculiare: se da un lato condiziona la possibilità di sviluppare contestualmente differenti attività di servizio, dall'altro ne può consentire un'amplificazione. In questo caso è un fattore che consente una variabilità basata sulla modificazione delle caratteristiche dell'offerta in momenti/periodi diversi.

2) Vincoli connessi alla disponibilità di risorse

La ricerca della creazione del valore per il cliente richiede la comprensione della natura e della potenzialità delle risorse di impresa; la sopravvivenza dell'impresa dipende da un'efficace gestione delle risorse aziendali, nelle componenti fisiche (strutture e attrezzature) e di personale (tempo e competenze), e

dei processi aziendali. In questa ottica, nell'ambito delle imprese di servizi, i vincoli di capacità produttiva esercitano un forte impatto sia sulle opportunità di generazione di valore per l'acquirente, sia sui costi aziendali. Il problema è quello di gestire il costo del potenziale connesso alle risorse aziendali, ossia di <<gestire con profitto le competenze, i processi e le capacità di un'organizzazione in modo da supportare adeguatamente l'indirizzo strategico del business>>. Il problema di fondo è quello di ottimizzare la gestione delle risorse in modo che si riducano gli sprechi. In relazione a tali obiettivi, è necessario affrontare il tema dei vincoli connessi alla capacità produttiva.

La loro natura cambia a seconda del servizio.

Si possono avere:

- Vincoli di capienza
- Vincoli di recettività
- Vincoli di tempo
- Vincoli di risorse umane

La gestione della capacità produttiva ha impatto, da un lato, sulla percezione della qualità del servizio da parte dei clienti; dall'altro sui costi dell'impresa.

Con riferimento al primo aspetto, in generale, situazioni di squilibrio tra presenze effettive e capacità produttiva considerata ottimale producono effetti differenti a seconda della tipologia dominante di benefici ricercati.

Nel caso il servizio risponda a benefici di carattere funzionale (core service), un numero di presenze superiore alla capacità produttiva, esercita un impatto negativo sulla qualità percepita e sulla soddisfazione del cliente in relazione a vari fattori: affollamento, code e tempi di attese elevati, riduzione della personalizzazione del servizio. Mentre un numero di clienti inferiore a quello corrispondente alla capacità produttiva ottimale produce un effetto positivo sui clienti, in ragione della maggiore comodità (si pensi ad un viaggio in treno), della riduzione dei tempi di attesa, della possibilità di avere una maggiore attenzione da parte del personale (si pensi allo sportello bancario).

Se il servizio risponde a benefici di carattere simbolico relazionale, la situazione tende ad invertirsi: prendendo ad esempio il caso dei locali notturni e di ristoranti di moda, l'affollamento diviene un fattore che incide positivamente sull'immagine; la presenza di poche persone ha un impatto negativo sulla qualità percepita dei locali. Passando a considerare l'aspetto dei costi, un dimensionamento della capacità produttiva non coerente con i flussi e le caratteristiche della domanda è all'origine di sprechi.

Il modello dell'utilizzo della capacità si basa sul principio che le risorse non utilizzate configurano l'esistenza di uno spreco; le diverse tipologie di spreco individuate sono:

- Spreco strutturale: deriva dal non corretto allineamento tra capacità reale e capacità necessaria per il soddisfacimento della domanda;
- Spreco predefinito: deriva dall'aver adottato una misura della capacità di base diversa da quella della capacità teorica;
- Spreco tecnico: è causato da variazioni nel mix di prodotti o da variazioni temporali dei flussi di domanda;
- Spreco di gestione: è determinato dalle politiche adottate (ad esempio chiusura di un locale il mezzogiorno nel fine settimana; sospensione dell'attività di call center in alcune festività).

Una cattiva gestione della capacità è inoltre legata ad un orientamento errato delle risorse aziendali rispetto alle specifiche esigenze dei diversi segmenti di mercato.

1.3.5 Il triangolo del servizio

Questa nuova concezione del rapporto azienda - cliente, è ben rappresentata nel "triangolo del servizio" di seguito riportato[1]:

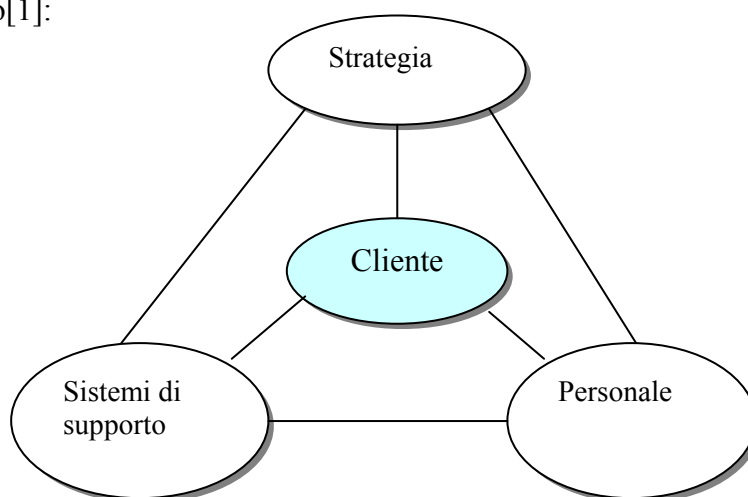


Figura 1.2 Il triangolo del servizio

Il centro è rappresentato dal cliente e le diverse attività dell'azienda sono preposte a servire il cliente e supportarlo nel miglior modo possibile. Un ruolo importante in questo senso lo svolge il personale, in quanto se la forza lavoro è ben motivata e addestrata, allora opererà al meglio per i propri clienti: come il management tratta il lavoratore, così il lavoratore tratterà il pubblico[1], ed è questo uno dei parametri di prestazione che più balza agli occhi dei consumatori e che fornisce loro la “percezione del servizio”. L'obiettivo è quindi quello di massimizzare i benefici che derivano dalle relazioni con il cliente, attraverso:

- Pratiche operative per identificare, attrarre e mantenere i migliori clienti
- Cultura aziendale basata su una strategia di lungo periodo di relazioni soddisfacenti con la clientela
- Insieme di moduli applicativi utilizzati per la conoscenza e la gestione del cliente
- Insieme integrato di strumenti destinati alla vendita, al marketing e alla “customer satisfaction”[12]

Gli obiettivi principali perseguiti quindi potrebbero essere così riassunti:

- Aumentare la conoscenza del cliente per poterlo servire meglio
- Consolidare e unificare la visione del cliente
- Coordinare le azioni fra le diverse strutture e/o aree di business e orientarle al cliente
- Modificare il tradizionale processo di marketing per focalizzarlo di più sul cliente

1.3.6 La qualità nei servizi

Nell'ambito dei servizi, la qualità non può che essere inquadrata in un'ottica cliente. Il servizio nasce come strumento di risposta ai suoi bisogni, desideri e attese. La prestazione è di qualità se il cliente può godere dei risultati. S'impone quindi all'organizzazione di non limitarsi a progettare il servizio sulla base delle proprie esigenze (i suoi problemi, i suoi vincoli, i suoi costi interni, le sue difficoltà), ma di impegnarsi sul cliente.

Quindi nell'ambito delle aziende di servizio bisogna procedere ad una valutazione dell'attività svolta relativamente alla qualità, quantità dei bisogni soddisfatti e quindi del grado di soddisfazione dei destinatari diretti. La valutazione riguarda aspetti attinenti al rapporto risultati ottenuti e risorse impiegate, la razionalità delle scelte effettuate tra le alternative potenzialmente disponibili ed il rispetto delle norme di comportamento.

La nozione di qualità non è facilmente definibile. In generale possiamo distinguere almeno due aspetti della qualità:

1. *la qualità estrinseca*, d'uso o funzionale;
2. *la qualità intrinseca*, progettuale o strumentale.

Ci si riferisce alla qualità estrinseca, d'uso o funzionale, qualora vengano considerate le caratteristiche che rendono un servizio idoneo a soddisfare i bisogni dei soggetti ai quali sono destinati.

Un servizio deve servire a soddisfare le esigenze, e quindi i bisogni, dei soggetti ai quali è destinato, pertanto dovrà essere tecnicamente utile. Oltre ad essere tecnicamente utile dovrà, però, soddisfare anche le esigenze esprimibili in termini di aspirazioni.

Gli indicatori che possono essere utilizzati per analizzare la qualità estrinseca sono volti a individuare:

- l'esistenza di un aggiornamento delle caratteristiche tecniche per mantenere il servizio idoneo a soddisfare i bisogni;
- la presenza di verifiche dei bisogni e delle aspirazioni della collettività destinataria dei servizi.

La qualità intrinseca, progettuale o strumentale, subentra qualora vengano considerate le caratteristiche che rendono un servizio conforme ad uno standard di riferimento. Lo standard di riferimento può essere rappresentato dai servizi offerti in passato, nel caso in cui si attui un confronto temporale. Sarà quindi importante che la qualità intrinseca si mantenga sempre conforme agli standard qualitativi prefissati o migliori nel tempo.

Gli indicatori relativi alla qualità progettuale sono volti ad individuare:

- il mantenimento dello standard tecnico uniforme nello spazio e nel tempo;
- l'esistenza di un servizio volto ad effettuare ispezioni e a verificare gli eventuali reclami.

A titolo esemplificativo possono essere considerati i seguenti parametri volti ad individuare la maggiore o minore qualità dei servizi:

- il rispetto di standard prefissati;
- la celerità nell'erogazione dei servizi;
- la tempestività nell'intervento in caso di problemi;
- l'aumento del grado di soddisfazione del cliente/utente.

1.3.6.1 Misurazione della produttività nei servizi

La valutazione del grado di soddisfazione dei destinatari di un servizio, si fonda sul *rapporto* tra *risultati ottenuti* e *risorse impiegate*. Questo rapporto non rappresenta altro che la *produttività* di un processo di trasformazione di risorse, ovvero il rapporto tra *output* e *input*, come in genere viene definito in campo manifatturiero.

Per misurare gli output e gli input impiegati nell'erogazione di un servizio la maggior parte delle organizzazioni fa uso di alcuni parametri quali “il numero di clienti serviti” come output, oppure “il tempo impiegato per servire i clienti” come input. Ma per la natura intangibile delle attività di servizio, una misurazione di tipo “monodimensionale”, basata cioè su un solo parametro di valutazione, non risulta molto appropriata; risulterebbe più idonea, pertanto, una analisi “multidimensionale”.

Quest'ultima si basa sul concetto che l'output intangibile di un'attività di servizio sia rappresentato dalla *qualità del servizio*, a sua volta funzione di cinque dimensioni quali [1]:

- 1) **Elementi tangibili:** aspetto della struttura fisica, delle attrezzature, del personale, nonché di altri fattori che agiscono sui sensi, quali rumore e temperatura.

- 2) **Affidabilità:** capacità di svolgere il servizio come promesso, in modo preciso e puntuale.
- 3) **Capacità di risposta:** prontezza nel fornire assistenza alla clientela.
- 4) **Rassicurazione:** competenza e cortesia degli addetti, corredata dalla loro capacità di trasmettere fiducia.
- 5) **Empatia:** attenzione premurosa e personalizzata.

L'input invece è rappresentato dal *livello di competenze* del personale dell'azienda di servizio, che si fonda invece sulla valutazione di sei parametri di riferimento quali [16]:

- 1) **Competenze tecniche e informatiche**
- 2) **Capacità di risolvere i problemi (problem solving)**
- 3) **Capacità di lavorare in gruppo**
- 4) **Qualità umane**
- 5) **Conoscenza del contesto operativo**
- 6) **Conoscenza delle logiche organizzative**

Alla qualità del servizio sarà assegnato un valore pari alla media dei punteggi assegnati alle cinque dimensioni caratteristiche di questo parametro, secondo una scala di importanza crescente che va da 1 a 7; analogamente alle sei dimensioni caratteristiche del livello di competenza del personale, verranno assegnati dei punteggi secondo una scala che va da 1 a 5, il cui valore medio rappresenterà il grado di incidenza che questo parametro può avere sulla *performance* del sistema. Il rapporto tra i due valori medi appena calcolati, darà come risultato *l'indice di produttività* del sistema di servizio considerato. Esso dovrà risultare maggiore o uguale di un indice di riferimento, ottenuto come rapporto tra il valore massimo che può assumere la qualità del servizio (7) e il valore massimo attribuibile al livello di competenze (5): tale indice rappresenterà *l'indice minimo di produttività* rispetto a cui confrontare quello reale, che sarà pari pertanto a $7/5 = 1,4$.

Se l'indice di produttività reale dovesse risultare inferiore a questo valore, vorrebbe dire che il processo di erogazione del servizio preso in considerazione, presenta dei potenziali "punti critici" che ne compromettono una realizzazione ottimale, sia dal punto di vista operativo che dal punto di vista qualitativo (aspetto quest'ultimo, di fondamentale importanza nell'ambito delle attività di servizio), sui quali quindi sarà necessario intervenire e porvi rimedio con soluzioni che interesseranno sia gli aspetti tecnici che quelli di natura strategica.

1.3.7 La dimensione tempo

Uno dei parametri più rappresentativi del livello qualitativo del servizio, è senza dubbio costituito dal "tempo". Nella realizzazione di tutti i servizi, dai più semplici ai più complessi, il cliente fornisce un proprio contributo mediante la donazione del proprio "tempo". La concretizzazione della soluzione rispetto alle attese è legata quindi certamente alle capacità dell'organizzazione erogatrice del servizio, ma anche all'utilizzo e al consumo del tempo del cliente. E oggi, più che in passato, il tempo ha un forte valore economico e sociale.

Raramente l'organizzazione conosce il servizio così come lo vive il cliente. In altri termini, a fronte di un servizio fornito da più soggetti che si interfacciano con il cliente in momenti diversi, vi è un'inadeguatezza di percezione.

Ogni soggetto dell'organizzazione ha una conoscenza parziale: solo il cliente ha la visione completa della dinamica del servizio[12]. Se l'organizzazione non si preoccupa di coordinare l'attività tra i diversi attori, c'è il rischio di produrre un mix di prestazioni, valide se considerate fuori dal contesto, ma non in armonia tra loro e che compromettono quindi la valutazione del cliente.

Le attese, le code, i tempi morti o gli spostamenti sono aspetti del servizio non sempre tenuti nella giusta considerazione. Eppure sono momenti in cui il contributo del cliente è notevole in quanto egli mette a disposizione il proprio tempo, che, se viene però perso in code e attese, diventa mal speso.

L'organizzazione deve quindi misurare il tempo globale del ciclo del servizio, sommando i tempi impegnati nello svolgimento delle attività a quelli delle attese, delle code, delle ripetizioni. E non può esimersi dal dare valore al tempo richiesto al cliente.

1.4 Classificazione dei servizi in funzione del grado di interazione col cliente

Un altro tipo di classificazione delle aziende di servizi, oltre a quelli visti in precedenza, può essere effettuata basandosi proprio sull'intensità del contatto del cliente nella creazione del servizio, definita come la percentuale di tempo durante la quale egli si trova all'interno del sistema, in rapporto al tempo totale necessario per garantire il servizio: maggiore è la percentuale di tempo di contatto, maggiore sarà il grado di interazione tra i due durante il processo di produzione [13]. In base quindi a queste considerazioni, si può definire:

- **Sistemi di servizio a elevata intensità di interazione**, in cui il cliente è direttamente coinvolto nel processo e per questo può influenzarne il tempo assorbito per l'esplicitazione della domanda, l'esatta natura del servizio o la qualità del servizio stesso.

- **Sistemi di servizio a bassa intensità di interazione**, in cui ovviamente la presenza fisica del consumatore nel sistema di erogazione e anche creazione del servizio, è molto limitata.

Risulta ben evidente come la prima categoria di servizi sia più difficile da governare e razionalizzare, in quanto il livello di partecipazione del cliente può variare molto.

Si riportano di seguito sinteticamente alcune differenze tra un sistema a bassa interazione ed uno ad alta interazione.

Decisione progettuale	Sistema a elevata intensità di interazione	Sistema a bassa intensità di interazione
Localizzazione struttura	Le operazioni devono aver luogo a contatto col cliente	Le operazioni vengono localizzate in prossimità delle risorse
Layout struttura	Funzionale alle aspettative psicofisiche del cliente	Orientato all'efficienza produttiva
Progettazione prodotto	Natura del servizio definita dal prodotto fisico e dall'ambiente di erogazione	Prodotto definito da minor numero di attributi
Progettazione processo	Le varie fasi hanno effetto diretto sul cliente	Cliente non coinvolto nelle varie fasi del processo
Schedulazione	Cliente coinvolto nel processo di schedulazione delle attività	Cliente interessato solo alla conclusione del processo
Pianificazione produzione	Livellamento produzione non possibile perché gli ordini non possono essere immagazzinati	Livellamento possibile
Competenze addetti	Forza lavoro diretta capace di interagire col pubblico	Forza lavoro diretta dotata solo di competenze tecniche
Controllo qualità	Standard di qualità variabili perché si identificano col punto di vista della clientela	Definiti e misurabili
Standard temporali	Flessibili a seconda delle esigenze cliente	Rigidi perché si lavora su articoli che sostituiscono il cliente
Sistema di retribuzione	Funzione del tempo perché l'output è variabile	Funzione dell'output perché esso è determinabile
Pianificazione capacità	Dimensionata sui picchi di domanda	Dimensionata sulla domanda media

Tabella 1.1 Differenze fondamentali tra sistemi a elevata e a bassa intensità di interazione

Si può osservare che su ogni decisione di progettazione, incide il fatto che il cliente presenzi o meno all'erogazione del servizio.

1.5 Servizi front-office e back-office

Per quel che concerne le caratteristiche delle tipologie di servizio con un diverso grado di contatto con la clientela, esistono in letteratura diverse filosofie di pensiero in merito. Chase, ad esempio, fa riferimento ad una classificazione in tre categorie: “servizio puro” (contatto elevato), “servizio misto” (contatto medio) e servizio “quasi manifatturiero” (contatto basso);

Fitzsimmons e Sullivan, invece organizzano la loro classificazione intorno al concetto di tipo di servizio rivolto al cliente: distinguono, quindi, il servizio “per il cliente” dal servizio “al cliente”;

Shostack, ancora, propone gli attributi di “complessità” e “divergenza”, riferendosi al numero di passi costituenti il processo di erogazione del servizio e al loro grado di complessità [2]. Più in generale, si potrebbe parlare di servizi *front-office* nel caso di attività svolte a diretto contatto col consumatore e *back-office*, i cui processi vengono eseguiti “dietro le quinte”.

Occorre precisare che questa distinzione non fa riferimento alla semplice presenza fisica o meno del cliente durante la distribuzione del servizio; infatti, se si analizza il caso dei trasporti pubblici, per esempio, il passeggero, seppur presente fisicamente, non risulta molto partecipe al servizio; quindi sarebbe più corretto fare una distinzione sul concetto di “contatto attivo” o “passivo”.

Quando il contatto del consumatore risulta passivo o inesistente, allora i processi del servizio possono essere progettati con gli stessi principi della produzione industriale, standardizzati e centralizzati, raggiungendo così gli obiettivi tipici dell'economia di scala, quali elevata efficienza, elevati volumi produttivi e riduzione dei costi; sono sempre più frequenti esempi di sistemi di servizi come ospedali o la carta stampata, che per il conseguimento di un certo livello di standardizzazione e centralizzazione delle attività di back-office, preferiscono eseguire le stesse, in strutture ubicate anche a parecchi chilometri di distanza dal centro di erogazione del servizio.

Quando invece il ruolo del consumatore risulta più attivo e partecipe, allora l'attenzione si poserà piuttosto sulle abilità dei dipendenti “in prima linea” e sulla loro capacità di soddisfare le aspettative dei clienti.

Per poter effettuare un confronto tra le caratteristiche peculiari di queste due tipologie di struttura organizzativa, si ricorre spesso all'analisi e confronto di alcuni parametri significativi quali[2]:

- *modalità di interazione* tra l'azienda erogatrice del servizio e il cliente
- *grado di clientelizzazione* (il livello di partecipazione del cliente alla realizzazione del servizio)

- *caratteristiche di progetto* del sistema di erogazione del servizio
- *capacità e priorità competitive* scelte dall'azienda
- *performance*

Si evidenzia che, per quanto riguarda la maggior parte dei parametri presi in considerazione, esistono differenze significative facilmente prevedibili tra le due tipologie di servizio, come nel caso delle “modalità di interazione”, del “livello di partecipazione del cliente” al processo, ecc. Su alcuni parametri invece, come quelli di “performance” si riscontrano dei risultati alquanto particolari. Infatti, contrariamente alle normali aspettative, gli investimenti di capitale sembrano costituire un fattore di maggiore peso nell’ambito delle organizzazioni front-office, mentre per la modalità back-office ciò che acquisisce maggiore importanza risulta essere la forza lavoro. Ma la ragione di tutto ciò risiede nel fatto che anche nei processi svolti “in prima linea”, è necessario investire ingenti quantità di capitale per provvedere ad esempio all’addestramento del personale, che costituisce la risorsa principale per il raggiungimento di una performance competitiva; così come i processi eseguiti “dietro le quinte”, devono necessariamente puntare su un elevato livello di forza lavoro (oltre che a ingenti investimenti di capitale) per poter adempiere alle proprie funzioni.

Un altro aspetto da evidenziare è che non è possibile in realtà effettuare una vera e propria distinzione tra le due tipologie di processo di servizio, ma che il “tradeoff” costituito da un elevato livello di qualità da un lato, e una maggiore efficienza del processo produttivo dall’altro, può essere risolto solo attraverso una concezione integrata dei due sistemi, considerando cioè il back-office come un sistema di supporto al front-office e non come una strategia impiegata unicamente con lo scopo di conseguire un abbattimento dei costi di produzione.

1.5.1 Disaccoppiamento o integrazione front e back office

Nel precedente paragrafo è stato già evidenziato come non sia possibile effettuare per certi aspetti una netta distinzione tra le due tipologie di processo. Ciò nonostante esiste la tendenza, in numerose imprese di servizi, a “disaccoppiare” il lavoro di back-office da quello a stretto contatto col cliente, poiché ciò comporta:

- Maggiore efficienza dovuta alla mancanza di interruzione delle attività
- Specializzazione delle attività
- Minore costo della manodopera
- Minore variabilità di volume

- Aumento dei volumi specifici: sorgono le economie di scala

Si riscontrano però anche i seguenti svantaggi:

- Maggior divario con il front-office
- Maggior *overlapping*, ossia sovrapposizione di attività
- Tempi di risposta maggiori
- Maggiore rigidità
- Impossibilità di avere un approccio per prodotto: si passa quindi ad un approccio per attività

Pertanto l'obiettivo dovrà essere quello di effettuare delle scelte riguardanti il giusto grado di disaccoppiamento e di centralizzazione, in funzione del tipo di strategia che l'azienda intende perseguire, tenendo ben presente l'impatto che tali decisioni avranno sulle prestazioni strategiche. Nella figura 1.4 sono rappresentate quattro tipologie di approccio strategico adottabili dalle aziende [6], in funzione dei costi (asse delle ascisse) e del livello di disaccoppiamento delle attività di back-office da quelle del front-line (asse delle ordinate). L'obiettivo sarà quello di illustrare come l'organizzazione del lavoro di back-office può supportare una data strategia, piuttosto che discutere della superiorità di un approccio rispetto ad un altro:

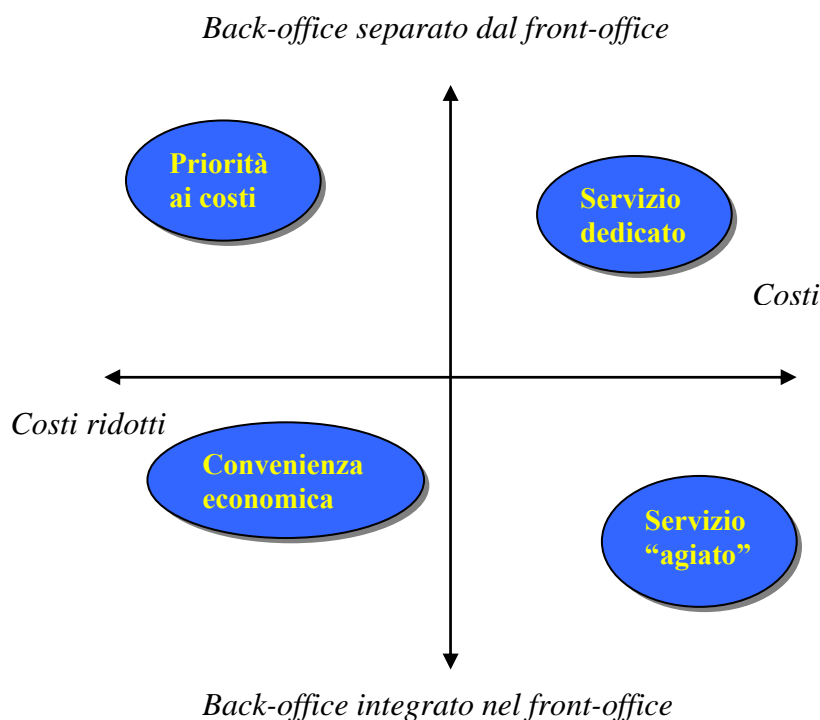


Figura 1.3 Le quattro tipologie di disaccoppiamento

1) Priorità ai costi : *disaccoppiamento elevato / costi ridotti*

Rappresenta la classica strategia usata soprattutto nel settore manifatturiero, per ridurre i costi di lavorazione, ma mentre in quest'ultimo settore l'obiettivo si raggiunge semplicemente suddividendo i lavori più complessi in operazioni elementari, standardizzate e eseguite da manodopera e tecnologia specializzate, aumentando il più possibile il volume produttivo, nel settore dei servizi, invece, è importante che le operazioni siano centralizzate, per ridurre la “varianza” del lavoro. Infatti, se i compiti venissero solo suddivisi e non centralizzati, le attività di back-office necessiterebbero di personale aggiuntivo per far fronte alle diverse esigenze della clientela, con la alternanza di alcuni giorni di produttività e altri di inattività; quando invece il lavoro di back-office viene centralizzato, i giorni di fermo relativi ad alcune operazioni, risultano equivalenti ai giorni di attività di altre, il che comporta un flusso di lavoro più prevedibile e una minore necessità di manodopera aggiuntiva[7].

Anche per quanto riguarda i tempi di risposta riscontriamo una certa lentezza, perché il tempo rilevante per il cliente è quello che va dalla richiesta del servizio alla sua erogazione, compresi quindi tempi di attesa e movimentazione, e sono proprio questi che si accentuano con la segregazione dei compiti; anche la stessa riduzione dei costi, inoltre, potrebbe essere raggiunta attraverso un coefficiente di utilizzazione della forza lavoro al 100%, cosa che è possibile forse nel settore manifatturiero, ma in quello dei servizi, esso risulta molto più variabile.

Risulta quindi evidente come questo tipo di approccio, che fa leva su elevata tecnologia, basso livello di manodopera e scarsa attenzione alle strategie di marketing, sia utile nel caso di richieste standard e non speciali che spesso non vengono prese neanche in considerazione, ma addirittura “scoraggiate” dai dipendenti del front-office. Inoltre la mancanza di conoscenza e partecipazione agli obiettivi delle proprie attività lavorative, da parte dei dipendenti del back-office, rende gli stessi meno motivati e quindi potenzialmente dannosi alla produttività della stessa azienda di cui fanno parte.

2) Convenienza economica : *disaccoppiamento minimo / costi ridotti*

Questo rappresenta invece l'approccio utilizzato da quelle imprese che cercano la riduzione dei costi attraverso innanzitutto una strategia a “chioschi” e cioè attraverso la dislocazione di tante strutture di servizio di piccole dimensioni e a produzione limitata, su tutto il territorio di competenza; inoltre la maggior parte di lavoro di back-office, deve essere svolta all'interno della struttura stessa, all'interno del front-office[8].

3) Servizio dedicato : *disaccoppiamento elevato / costi elevati*

Rispetto all'approccio contrapposto (Priorità ai costi), l'elevato grado di centralizzazione delle funzioni di back-office in questo caso, è accompagnato da una stretta relazione con il personale dell'altro fronte, in maniera tale che i dipendenti "dietro le quinte" non siano più costretti a lavorare per clienti che non hanno mai neppure visto, ma anzi, che si sentano più coinvolti e responsabili; tutto ciò conferisce una maggiore varietà e flessibilità nell'offerta del servizio e conseguentemente maggiori costi.

4) **Servizio "agiato"** : *disaccoppiamento minimo / costi elevati*

Rappresenta quella che Porter chiamava "strategia di differenziazione" [5], corrispondente all'erogazione di un eccellente livello di servizio personalizzato, con elevata partecipazione del consumatore che, in casi estremi, risulta familiare al personale in maniera tale da conoscere in anticipo i suoi desideri e le sue aspettative; tutto questo comporta elevati costi e un disaccoppiamento del lavoro di back-office, realizzato solo nel caso di considerevoli vantaggi derivabili da economie di scala.

È evidente quindi come i lavoratori a stretto contatto con la clientela in questo tipo di approccio, possano recepire delle informazioni sulle aspettative dei consumatori[8], che risulteranno utili per il conseguimento di ottimi risultati anche nel lavoro di back-office.

Si può concludere dicendo che la scelta quindi di come organizzare il lavoro di back-office nelle attività di servizio, non deve basarsi solo su una decisione tattica fatta nell'interesse dell'efficienza, ma in accordo con il tipo di strategia messa in atto dall'azienda; se è vero che attraverso il disaccoppiamento si possono ottenere costi più bassi, elevate velocità e standard di qualità, è anche vero che se questa scelta va in contrasto con le decisioni strategiche e con la funzione ugualmente importante del front-office, tutti gli obiettivi suddetti difficilmente potranno essere raggiunti [3].

1.6 L'operations management nei servizi

Servire *bene* un cliente, come è stato evidenziato, significa operare tempestivamente, con qualità eccezionale e al minor costo possibile. Ciò si traduce nella progettazione di processi *efficienti* ed *efficaci*. A efficienza ed efficacia si collega il concetto di *valore*, che può essere metaforicamente definito come il rapporto qualità-prezzo: se si può fornire a un cliente un'automobile migliore senza variare il prezzo, il valore aumenta; se si riesce a dare al cliente un'auto migliore a un prezzo inferiore, il valore si impenna. Uno degli obiettivi del management è quello di conseguire elevati livelli di valore per l'azienda.

Per **operations management (OM)**, si intende il processo di progettazione, realizzazione e miglioramento dei sistemi d'impresa che creano prodotti o erogano servizi. Come il marketing e la finanza, l'OM è una

funzione d'impresa con precise responsabilità gestionali. Produrre un bene o fornire un servizio richiede una complessa serie di processi di trasformazione: l'OM è deputata al coordinamento di tutti questi processi distinti, nel modo più efficace possibile. Nell'ambito della funzione di operations, le decisioni gestionali possono essere suddivise in tre macrocategorie:

- ❑ decisioni strategiche (di lungo periodo);
- ❑ decisioni tattiche (di medio periodo);
- ❑ decisioni di pianificazione operativa e di controllo (di breve periodo).

Le scelte strategiche di OM generano importanti ricadute sulla capacità aziendale di soddisfare i bisogni dei consumatori nel lungo periodo. Perciò, perché l'impresa abbia successo, queste decisioni devono essere coerenti con la strategia aziendale più complessiva. Le decisioni assunte a livello strategico costituiscono le condizioni invariabili e i vincoli operativi entro i quali l'impresa potrà agire nel breve-medio periodo. Al secondo livello del processo decisionale si colloca la pianificazione tattica che, tra i suoi compiti chiave, ha l'efficiente organizzazione della manodopera e dell'acquisizione dei materiali, nel quadro precedentemente delineato dalle decisioni strategiche. Le decisioni tattiche, a loro volta, vanno a costituire i vincoli operativi entro i quali l'impresa potrà assumere le decisioni di pianificazione operativa e controllo.

Le scelte gestionali di pianificazione operativa e di controllo sono di più ridotta portata e più a breve termine, incentrate su vincoli e criticità nel breve periodo.

1.6.1 Sviluppo storico dell'OM

A cavallo degli anni Cinquanta e Sessanta tra i cultori della materia si comincia ad affrontare il tema specifico dell'OM e a rilevare le comunanze dei problemi affrontati da tutti i sistemi di produzione; si comincia così a guardare alle dinamiche delle operations e della produzione come a un sistema e si evidenzia la necessità di applicare approcci e metodi quali la "Teoria delle Code", la simulazione e la programmazione lineare, da allora divenuti temi di studio tipici nel campo delle operations [10].

Gli anni Ottanta vedono una rivoluzione nelle filosofie gestionali e nelle tecnologie mediante le quali ha luogo il processo produttivo. La più radicale innovazione nella produzione industriale è il modello *just-in-time* (JIT). Di concezione giapponese, il JIT si compone di un insieme di attività integrate volte a conseguire elevati volumi di produzione, detenendo quantitativi minimi di scorte, le quali giungono sul luogo di lavorazione soltanto al momento di effettiva necessità. Il JIT è oggi una pietra angolare delle logiche produttive di numerose imprese[1].

Certamente non furono i giapponesi i primi a ideare un sistema di produzione efficiente e altamente integrato. Già nel 1913 Henry Ford aveva ideato la catena di montaggio per costruire l'autovettura modello T. Il sistema sviluppato da Ford era soggetto ai soli vincoli delle capacità delle maestranze e delle tecnologie esistenti. Occorreva una perfetta programmazione e organizzazione perché operai e macchine fossero sempre saturi e operassero in modo sincronizzato con l'incessante arrivo dei materiali. Prodotto, processo, materiali, logistica e personale erano ben integrati e bilanciati sin dalla fase di progettazione dell'impianto e delle sue operations [11].

Tra gli anni Settanta e Ottanta si fa strada il "paradigma della strategia produttiva" formulato da alcuni studiosi della Harvard Business School [1]. Questo approccio esorta i dirigenti industriali a impiegare i punti di forza delle proprie fabbriche come arma competitiva strategica. Perno di questo modello è la nozione di fabbrica focalizzata e di trade-off tra performance produttive: poiché uno stabilimento non può raggiungere risultati eccellenti in tutti gli indicatori di prestazione, il suo management deve adottare una strategia di focalizzazione in grado di perseguire l'eccellenza su un limitato set di performance. Per fare ciò bisogna progettare e gestire il sistema produttivo sapendo scegliere fra i molteplici trade-off prestazionali (costi ridotti, qualità elevata, flessibilità, ecc.).

Tra la fine degli anni Ottanta e i primi Novanta, il fuoco si sposta sulla gestione della qualità totale (*total quality management*, TQM). Oggi sono le Norme ISO 9000, create dall'International Organization for Standardization, significativamente riviste nella nuova serie ISO 9000-2000, a svolgere il ruolo chiave nella definizione dei parametri di qualità per i produttori di tutto il mondo.

La recessione globale degli anni Novanta ha indotto molte società ad assumere configurazioni più snelle per rimanere competitive, e dunque a perseguire l'innovazione nei processi di gestione delle operations. Il messaggio di fondo della reingegnerizzazione dei processi aziendali (*business process reengineering*, BPR) è quello di effettuare cambiamenti radicali, in contrapposizione con l'approccio graduale suggerito dal TQM. Ciò si ottiene esaminando con occhi nuovi lo svolgimento dei processi aziendali, eliminando le fasi non a "valore aggiunto" e informatizzando le fasi rimanenti, per ottenere i risultati desiderati.

Alla fine degli anni Novanta l'attenzione si è spostata sul "supply chain management" e cioè su un approccio sistemico al governo dei flussi (di informazioni, materiali e servizi) che dai fornitori di materie prime, passando per le fabbriche e i magazzini, giungono al consumatore finale. Le nuove tendenze, quali l'outsourcing e la mass customization, stanno spingendo le aziende a trovare modi flessibili per soddisfare la domanda dei consumatori. Il punto centrale consiste nell'ottimizzazione delle attività "core" per garantire la massima prontezza di risposta alle variazioni nelle attese dei clienti.

Ma la fine degli anni '90 si segnala anche per lo sviluppo di Internet. L'espressione *commercio elettronico* si riferisce all'impiego del World Wide Web come elemento portante dell'attività imprenditoriale. Queste nuove modalità operative sono connotate dall'espressione *e-ops* (operations elettroniche)[1].

1.6.2 Operations management e tendenze attuali

Si è discusso finora dei temi salienti con cui oggi si confrontano i responsabili delle operations, che si possono riassumere nei seguenti punti:

- ❑ Integrare efficacemente le operations risultanti da operazioni di fusione e acquisizione. Oggigiorno le fusioni di grandi imprese sono alquanto frequenti. Spesso queste fusioni sembrano promettere molto in termini di economie di scala e di efficienza operativa..
- ❑ Sviluppare una supply chain flessibile per consentire la personalizzazione di massa di prodotti e servizi. Virtualmente tutte le aziende vanno ampliando le proprie linee di prodotto per offrire quella varietà di scelte che i consumatori richiedono.
- ❑ Gestire le reti globali di fornitura, produttive e distributive. L'implementazione di sistemi Global ERP (*Global Enterprise Resources Planning*) impiegati per la gestione delle risorse produttive, oggi piuttosto diffusa nelle grandi aziende, impone ai manager di sfruttare al meglio questa cospicua mole di informazioni. Ciò richiede tra l'altro, di scegliere con cura quando ricorrere a modalità di controllo accentrate e quando lasciare maggior autonomia e decentramento.
- ❑ Standardizzare e omologare le forniture. In molte occasioni esistono più fornitori in grado di offrire i beni richiesti da un'azienda. Si analizzi il caso dei fornitori di acciaio per uso automobilistico, per esempio [1]. In passato produttori di vetture e fornitori di acciaio, privilegiavano contratti a lungo termine, più idonei a giustificare investimenti in sistemi informativi integrati e in altri processi. L'attuale sviluppo della comunicazione via Internet e i progressi tecnologici dell'industria pesante rendono invece oggi meno dispendioso il passaggio da un fornitore all'altro. Per definire questa possibilità di scegliere alla pari fra diversi operatori, i fornitori usano l'espressione informatica *plug compatible*. Ora i produttori di acciaio possono perseguire accordi di partnership in forme alquanto più aggressive. Le aste on-line negli scambi *business to business* hanno mutato in modo clamoroso le dinamiche relazionali fra imprese.
- ❑ Arrivare alla "fabbrica dei servizi". Con questa locuzione si definisce la crescente spinta verso la creazione di servizi personalizzati per ciascun consumatore, nonostante un'azienda possa

contare milioni di clienti, mantenendo comunque spiccate caratteristiche di efficienza operativa e impiegando stabilimenti capaci di elevati volumi produttivi.

- ❑ Potenziare i servizi a valore aggiunto. Non basta più consegnare prodotti validi. I clienti aziendali vogliono rimanere al passo con i progressi della produzione, ricevere anticipatamente notizie dei cambiamenti apportati ai modelli, ottenere supporto per implementare le modifiche introdotte, fruire di unità di assistenza ben qualificate.
- ❑ Rendere efficiente l'uso della tecnologia Internet. La quasi totalità delle aziende più importanti utilizza Internet nell'ambito dei propri processi produttivi. Integrare progressivamente il Web in ciascuna fase della produzione richiede una struttura complessiva coerente, in grado di coordinare armonicamente più attività, potenzialmente osservabili come portafoglio di Web.
- ❑ Ottenere servizi validi dalle imprese di servizi. Tradizionalmente la produzione di beni fisici si impernia sull'uso efficiente delle risorse per creare il prodotto finale. Anche la produzione di servizi non può prescindere dall'efficienza, ma vi deve aggiungere la capacità di gestire l'apporto e l'esperienza del consumatore nel momento in cui si esplica la relazione di servizio. Il diverso orientamento tra la produzione di beni e quella di servizi emerge a pieno quando si tratta di intensificare tale relazione, operazione che può essere inefficiente dal punto di vista della mera allocazione delle risorse, ma fattore determinante della qualità del servizio percepita dal consumatore.

Obiettivo finale risulta essere, quindi, la realizzazione di processi produttivi, anche nell'ambito delle aziende di servizi, che siano al contempo efficienti ed efficaci. Tutto ciò si traduce nel bilanciamento ottimale della capacità produttiva aziendale.

CAPITOLO II

IL BILANCIO DELLA CAPACITA' PRODUTTIVA IN BACK OFFICE: LA SCHEDULAZIONE DEL PERSONALE

Una delle principali operation nelle aziende di servizi è la gestione della forza lavoro. I due scopi principali sono:

- 1. Individuare il modo per migliorare la produttività e la qualità dei servizi, considerato che rappresenta la principale risorsa da manovrare.*
- 2. Gestire le risorse umane ed i rispettivi impegni lavorativi.*

Il principale strumento disponibile per queste operazioni è la schedulazione del lavoro.

Poiché i programmi di lavoro determinano il numero di risorse umane presenti durante differenti ore del giorno e differenti giorni della settimana, essi hanno importanti implicazioni sui costi operativi e sul servizio ai clienti. Se il numero di impiegati schedulati non è sufficiente a soddisfare la domanda del cliente, allora il sistema operativo può scegliere di lasciare il livello di servizio basso o può incrementare il livello di lavoro attraverso altri mezzi.

Allo stesso modo se il numero di impiegati sono in eccesso rispetto alla richiesta della domanda, allora si va incontro a costi di lavoro inattivo. La ricerca di tecniche di schedulazioni del lavoro si è focalizzata su politiche di staffing e scheduling, sul numero di lavoratori part time, sulle politiche di straordinario, sulla valutazione del numero di periodi di lavoro in un turno di lavoro giornaliero ed il numero di giorni di lavoro in una settimana, sulla sovrapposizione di turni da un giorno di lavoro al successivo, sulla disponibilità limitata di impiegati, ecc..

La schedulazione del personale consiste nell'organizzare gli orari di lavoro dei dipendenti di un'azienda al fine di soddisfare la domanda di beni o servizi che essa offre.

2.1. Labor Scheduling

Nel capitolo precedente è stata sottolineata la centralità delle risorse umane nelle aziende di servizi sia per quel che concerne l'impatto sui clienti e sulla relativa soddisfazione, sia perché rappresentano una sostanziosa aliquota dei costi sostenuti dalle aziende.

Una corretta gestione delle risorse umane è quindi di fondamentale importanza influenzando sia l'efficienza sia l'efficacia dei processi nelle aziende di servizi.

Una delle principali problematiche connesse alle risorse umane operanti in un'azienda di servizi è la gestione della forza lavoro, i cui scopi principali sono, in primo luogo, individuare modi per migliorare la produttività del personale, e, infine, gestire le risorse umane e i rispettivi impegni lavorativi.

Il principale strumento adoperato per la realizzazione di tali obiettivi è la schedulazione del lavoro.

Per comprendere bene il funzionamento della procedura di schedulazione del personale risulta utile considerare una specifica area del servizio e procedere ad un'analisi particolareggiata di ogni fase cercando di approfondire tutti gli aspetti relativi ad essa.

Lo scheduling dello staff costituisce una notevole proporzione dei costi negli alberghi, ristoranti fast food, complessi turistici impegnati a garantire un certo standard di servizio necessario per accontentare le richieste dei clienti.

In tali casi i costi dello staff rappresentano la spesa maggiore, pari all'incirca al 30% delle spese totali. Per questa ragione una riduzione anche solo dell'1% di tale spesa consente di realizzare un significativo risparmio.

In ogni dato momento di un orizzonte temporale di riferimento avere un numero insufficiente di impiegati, o averne un numero sufficiente privo della necessaria competenza, può comportare una perdita di vendite causata da un inadeguato livello di servizio per i consumatori e da un sovraccarico di lavoro per i dipendenti. D'altra parte un sovradimensionamento dello staff in servizio crea dei costi di lavoro inattivo e una demoralizzazione del personale, costretto a lavorare meno di quanto desiderato.

Il rostering è definito come l'atto di bilanciare le richieste dei consumatori con la domanda di impiegati e la redditività. Molte tipologie di aziende di servizi sono caratterizzate dal fatto che il costo delle risorse umane rappresenta una grande porzione dei costi sotto il controllo manageriale.

Controllare i costi attraverso lo scheduling può rappresentare una vera sfida a causa della necessità di tener conto di molteplici aspetti quali: desideri del personale, regolamentazioni governative, politiche della compagnia, obbligazioni contrattuali, ecc.

Le quattro fasi di cui si compone la procedura sono:

- 1) Previsione della richiesta del servizio da parte di consumatori**
- 2) Traduzione della richiesta prevista in domanda di impiegati**
- 3) Scheduling degli impiegati**
- 4) Adattamento dello scheduling effettuato alla reale richiesta del servizio**

Il primo passo della procedura consiste nel realizzare una previsione circa la richiesta del servizio da parte dei consumatori in un dato periodo di tempo, e circa la sua evoluzione nell'arco dell'orizzonte temporale di riferimento scelto.

La seconda fase utilizza come input il primo step calcolando, per ogni intervallo temporale, il numero di impiegati necessari a soddisfare la richiesta del servizio prevista.

La terza effettua lo sviluppo della schedulazione della forza lavoro tenendo conto delle competenze e dei desideri dei lavoratori, delle regole stabilite dalla compagnia e dai contratti lavorativi, delle norme legislative, ecc.

L'ultima fase, infine, confronta la schedulazione ottenuta con le prime tre fasi e la realtà operativa operando gli opportuni aggiustamenti necessari per fornire il servizio effettivo[14].

2.2. PRIMA FASE: DETERMINAZIONE DELLA RICHIESTA DEL SERVIZIO

La prima fase della procedura di scheduling si compone di otto passi fondamentali, ossia:

- 1) **determinare la natura del lavoro**
- 2) **identificare i fattori (“*labor drivers*”)che influenzano il lavoro**
- 3) **determinare se i fattori chiave del lavoro variano nel breve periodo**
- 4) **determinare l’intervallo di tempo opportuno per tracciare la variazione dei labor drivers**
- 5) **prevedere la variazione temporale dei labor drivers**
- 6) **ridurre con tecniche di smoothing la variazione casuale dei drivers**
- 7) **controllare, attraverso misurazioni e rilevamenti, l’accuratezza della previsione**
- 8) **definire il periodo di tempo durante il quale il lavoro può essere realmente svolto(“*finestra*”)**

I primi quattro passi della procedura devono essere ripetuti periodicamente, ad esempio ogni sei mesi, i restanti quattro invece vanno riprodotti ogni qualvolta si sviluppa una nuova schedulazione[15].

2.2.1. Passo 1: determinare la natura del lavoro

Lo scheduling del lavoro fatto dai managers del settore dell’ospitalità richiede di classificare il lavoro in due tipologie:

- lavoro controllabile
- lavoro non controllabile

Con il termine “lavoro controllabile” si intende un tipo di attività caratterizzata da flessibilità di esecuzione nel tempo a causa di nessuna interazione esistente tra il cliente e la risorsa umana impiegata dall’azienda; in caso contrario si parla di lavoro “non controllabile”[16]. Per il lavoro controllabile bisogna definire un intervallo temporale all’interno del quale deve essere svolta l’attività, esso viene chiamato “finestra” e può avere una lunghezza variabile in relazione al tipo di mansione considerata.

La schedulazione riguarda sia il lavoro controllabile sia quello non controllabile.

Per il primo risulta necessario determinare i tempi di esecuzione al più presto e al più tardi, ossia i limiti dell’intervallo di esecuzione.

Per il secondo bisogna, invece, prevedere i volumi di lavoro generati in tutti gli intervalli di pianificazione in cui è stato suddiviso l’orizzonte temporale di riferimento.

I prossimi sei step della prima fase riguardano entrambi i tipi di lavoro, l’ultimo step, invece, è specifico per il lavoro controllabile.

2.2.2. Passo 2: identificare i fattori che influenzano il lavoro

Tale passo consiste nell'identificare i fattori che influenzano il numero e le competenze degli impiegati addetti alla realizzazione e alla distribuzione del servizio. I labor drivers sono determinati trovando gli elementi che incidono sul tempo necessario per lo svolgimento di un determinato compito.

Una volta stabilito quali sono tali caratteristiche, tramite la tecnica di regressione multipla, si quantizza per ognuna l'influenza precisa sull'esecuzione dell'attività in questione. È importante che siano scelti, per ogni mansione, drivers indipendenti, ossia non legati fra loro. Per identificare labor driver indipendenti è possibile utilizzare una matrice di correlazione. Essa identifica la correlazione tra tutte le coppie di potenziali drivers e può essere generata automaticamente mediante una funzione incorporata in molti tabulatori software.

I labor driver che presentano un valore di correlazione inferiore a 0.5 possono essere considerati indipendenti.

2.2.3. Passo 3: determinare se i fattori chiave del lavoro variano nel breve periodo

Il terzo step consiste nel determinare un eventuale variazione dei labor drivers nel breve periodo.

I drivers possono essere di due tipi:

- drivers tempo varianti
- drivers tempo invarianti

Il primo subirà delle variazioni in tutti i periodi dell'orizzonte di riferimento, mentre il secondo rimarrà pressoché costante, al più presenterà valori tutti appartenenti ad una determinata fascia.

Nel lungo termine tutti i drivers risultano tempo varianti. Ciò è dovuto, solitamente, ad un aumento di efficienza nello svolgimento delle attività realizzato attraverso miglioramenti tecnologici.

Tuttavia tale aspetto non inficia l'analisi in questione che riguarda solo l'orizzonte di scheduling, ossia un periodo di tempo che varia da una settimana a tre mesi.

Per distinguere i drivers tempo varianti da quelli tempo invarianti è possibile utilizzare due metodi:

- 1) diagrammi che mettono in relazione il driver al tempo
- 2) analisi di correlazione

L'uso di un diagramma, che presenta sull'asse delle ordinate i valori del driver e su quello delle ascisse il tempo, permette di visualizzare l'andamento di una curva rappresentativa del driver in questione. Se tale curva mostra cambiamenti ciclici nel tempo allora il driver è tempo variante.

Se essa rimane relativamente costante o mostra una variazione casuale (ossia all'interno di una determinata fascia) allora si tratta di un driver tempo invariante.

La figura 1[15] mostra il plot di due labor drivers in un periodo di pianificazione giornaliero.

Il driver B sembra avere un andamento stabile durante il giorno, mentre quello A sembra variare in ogni intervallo. In base a tale figura si potrebbe affermare che A è un driver tempo variante e B è un driver tempo invariante. Il secondo metodo prevede la determinazione di un indice di correlazione ottenuto dal confronto tra i dati di un periodo con quelli del periodo precedente. Un esempio di tale procedimento è mostrato in figura 2: in un intervallo di 2 ore diviso in 8 periodi di 15 minuti l'uno è possibile osservare i seguenti valori:

\$13.28; \$15.41; \$14.62; \$14.93; \$17.03; \$15.56; \$12.10; \$15.51;

Utilizzando una funzione contenuta semplicemente in excel (“=CORREL(B1;H1;B2;H2)”) è possibile confrontare tali valori con quelli precedenti ottenendo come valore di correlazione 0.201.

Quando il valore di correlazione è minore di 0.5 allora il driver è tempo invariante, quando è maggiore o uguale di 0.5 il driver è tempo variante. Nel caso in esame il driver è tempo invariante.

Ripetendo il procedimento con i drivers A e B nell'esempio di figura 2[15] abbiamo rispettivamente i seguenti valori di correlazione: 0.927 e 0.101 ossia A è tempo variante e B è tempo invariante.

Come si può vedere l'esito dei due procedimenti illustrati è concorde.

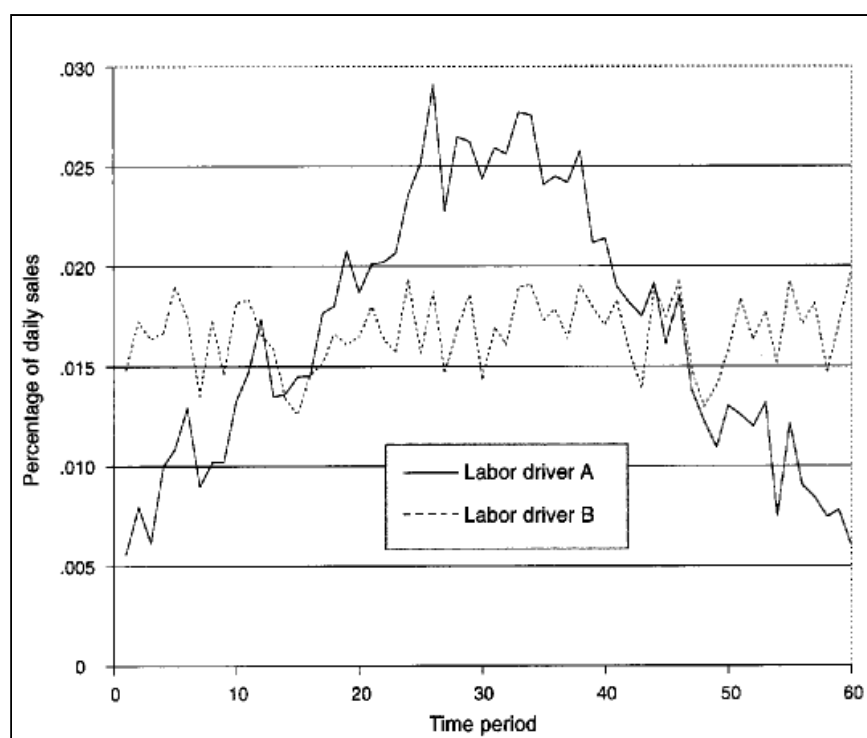


Figura 2.1: Plot di due labor driver in un periodo di pianificazione giornaliero

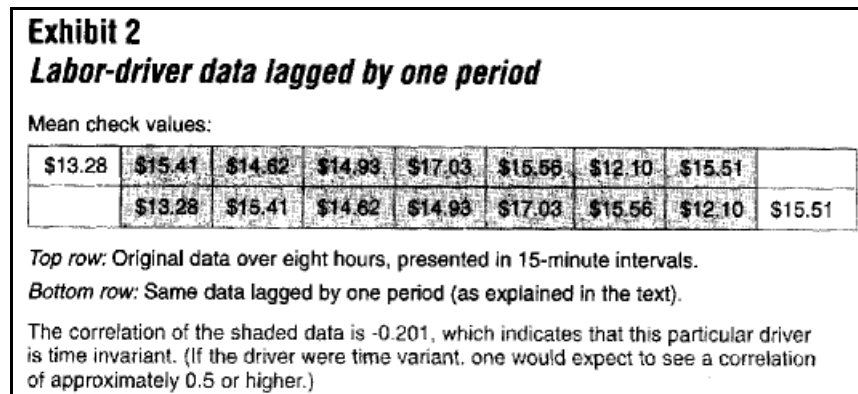


Figura 2.2: Uso dell'analisi di correlazione

2.2.4. Passo 4: determinare l'intervallo di tempo opportuno per tracciare la variazione dei labor drivers

Una volta determinati i driver tempo varianti e tempo invarianti, è necessario identificare un intervallo di tempo opportuno in cui tracciarli.

Per comprendere bene come effettuare tale scelta è utile prendere in considerazione la figura numero 3[15].

In essa sono rappresentati 4 set di dati relativi a un solo labor driver, plottati in quattro differenti intervalli di tempo:

15 min; 45 min; 150 min; 450 min.

La curva corrispondente all'intervallo di 450 min si mostra particolarmente piatta suggerendo, erroneamente, che il driver in questione è tempo invariante.

Gli intervalli di 150 min e 45 min generano delle curve che sottostimano o sovrastimano la curva creata dall'intervallo di 15 min che si mostra quindi il più adatto a rappresentare la variazione del driver.

Un intervallo di 15 min è comunemente usato nello sviluppo della schedulazione del lavoro anche in relazione al fatto che le pause degli impiegati, spesso, hanno proprio tale durata.

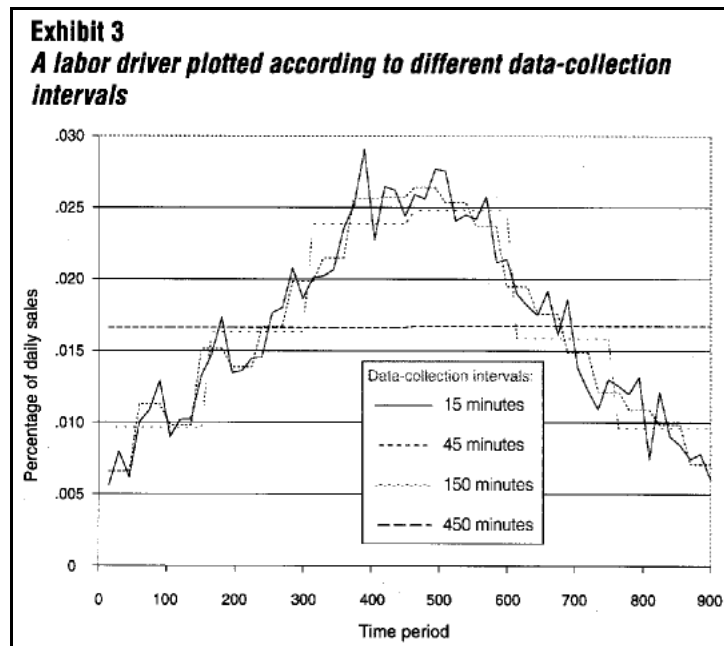


Figura 2.3: Scelta di un intervallo di tempo opportuno per tracciare variabilità di un Driver

2.2.5. Passo 5: prevedere la variazione temporale dei labor drivers

Il quinto passo consiste nella determinazione della variazione dei drivers tempo varianti in ogni periodo di tempo in cui è diviso l'intervallo di schedulazione.

Tale previsione può essere realizzata utilizzando due differenti approcci :

- 1) approccio indipendente
- 2) approccio di aggregazione e disaggregazione

L'approccio indipendente sviluppa una previsione, per ognuno dei periodi dell'orizzonte di riferimento, basandosi sull'assunzione che ogni periodo è indipendente da ogni altro.

L'approccio di aggregazione e disaggregazione richiede per prima cosa la raccolta di dati storici relativi a ogni periodo in cui è stato suddiviso un determinato intervallo temporale, fatto ciò, attraverso la combinazione di dati raccolti il metodo consente di realizzare facilmente una previsione per un qualsiasi altro intervallo di tempo, più lungo o più corto di quello iniziale. Per la sua applicazione si richiede però la consistenza nei dati.

La figura 4[15] mostra un esempio di applicazione di tale metodo. L'intervallo in cui sono collezionati i dati è di 4 settimane suddiviso in periodi di 1 settimana ciascuno. I dati sono relativi alle vendite totali giornaliere espresse come quota delle vendite totali della settimana (labor driver).

Dal diagramma si evince che il giorno della settimana in cui si realizzano volumi di vendita maggiori è il sabato, seguito nell'ordine dal venerdì, giovedì, domenica, lunedì, martedì, mercoledì.

Partendo da tale collezione di dati si potrebbe ricavare, con l'approccio di aggregazione e disaggregazione, una previsione in un qualsiasi nuovo intervallo di riferimento. Naturalmente ciò risulta possibile solo se esiste consistenza nei dati collezionati (ossia se essi rappresentano variabili perfettamente correlate). Attraverso la misurazione dell'indice di correlazione, riportata in figura 5[15], si accerta tale consistenza (evidenziata da un indice di correlazione ovunque maggiore di 0.94 quindi molto prossimo all'unità) e dunque la possibilità di usare l'approccio anzidetto.

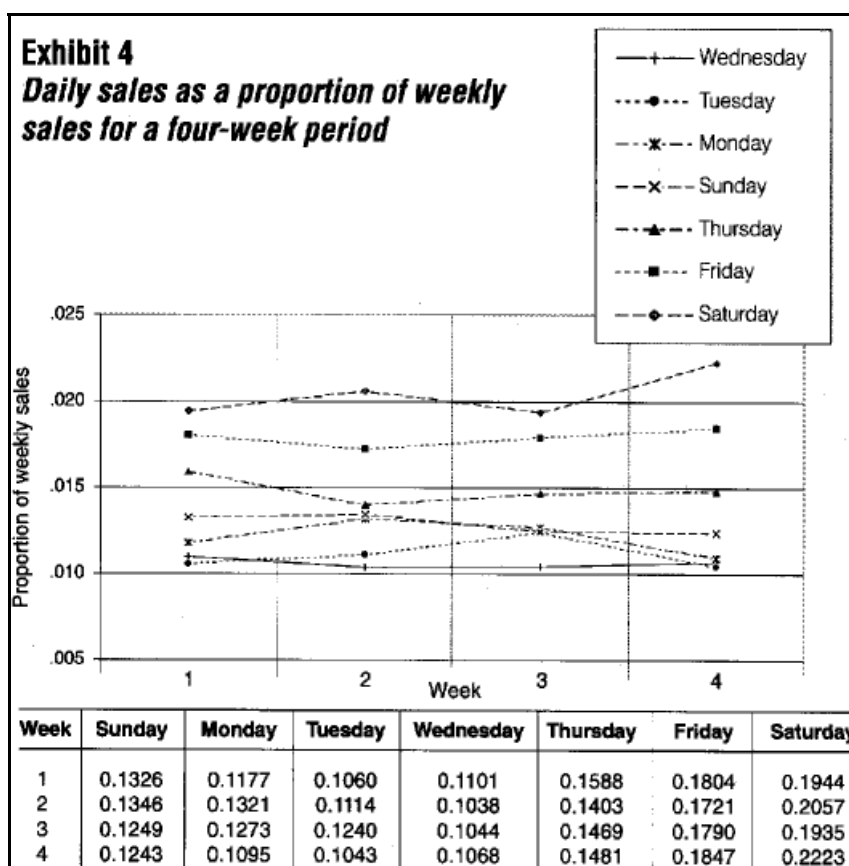


Figura 2.4: Esempio di approccio di aggregazione e disaggregazione

Exhibit 5 <i>Correlations between weeks for the data in Exhibit 4</i>				
	<i>Week 1</i>	<i>Week 2</i>	<i>Week 3</i>	<i>Week 4</i>
<i>Week 1</i>	1			
<i>Week 2</i>	0.9454	1		
<i>Week 3</i>	0.9563	0.9664	1	
<i>Week 4</i>	0.9760	0.9701	0.9679	1

Figura 2.5: Uso della correlazione per determinare la coerenza nei dati

2.2.6. Passo 6: ridurre la variazione casuale con tecniche di smoothing

Parte della variazione di un labor driver è prevedibile, parte invece è casuale quindi imprevedibile. Per un'efficace implementazione della procedura di scheduling è necessario cercare di ridurre la seconda. Solitamente essa viene messa in evidenza dalla presenza di picchi e vallate all'interno dei diagrammi che relazionano il labor driver al tempo. La figura 6[15] illustra una previsione di vendite per un particolare giorno della settimana (es:lunedì) in un orizzonte temporale di 4 settimane diviso in periodi da 15 minuti ciascuno. Nell'andamento della curva viene riscontrata la presenza di alcuni "denti". La ragione più probabile di tale presenza risiede spesso in un'insufficienza numerica di dati rilevati, essa genera una variazione casuale intorno al vero valore della domanda in quel periodo. Per eliminare parte della variazione casuale si possono utilizzare tecniche di smoothing che consentono un appianamento della curva facendo una media tra le previsioni di quel periodo e di quelli ad esso adiacenti, ossia quello precedente e quello successivo. Il pericolo nell'uso di previsioni realizzate attraverso lo smoothing si verifica quando i picchi e le vallate sono causati da fenomeni reali piuttosto che da difetti nei dati. In figura 7[15] vengono riprodotti i risultati ottenuti dall'applicazione dello smoothing al diagramma in figura 6.

Exhibit 8

Forecast of demand, by 15-minute periods

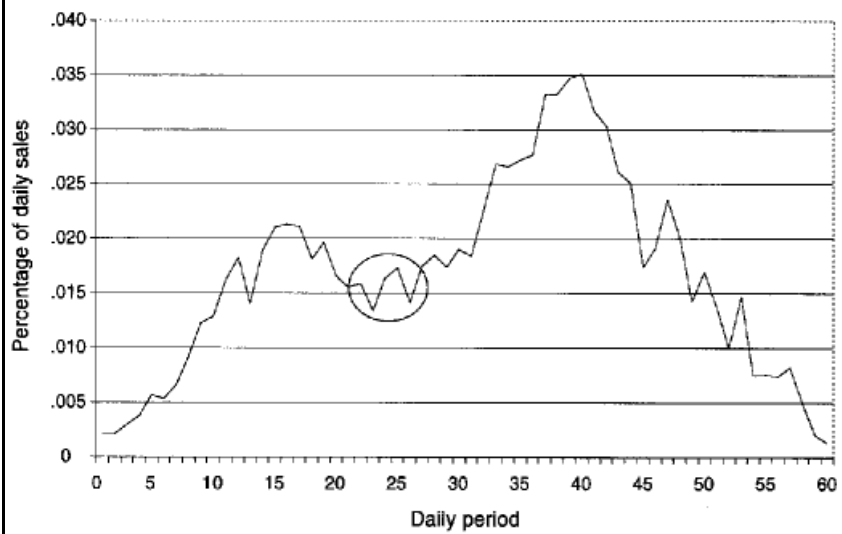


Figura 2.6: Uso delle tecniche di smoothing

Exhibit 9

Comparing unsmoothed and smoothed demand forecasts

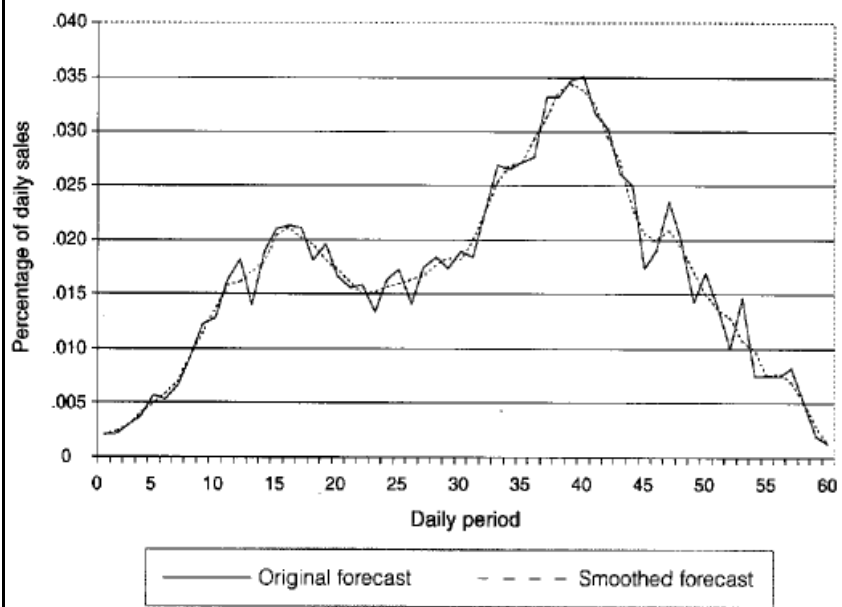


Figura 2.7: Risultati prodotti dallo smoothing

2.2.7. Passo 7: controllare, attraverso misurazioni e rilevamenti, l'accuratezza della previsione

Raramente una previsione risulta perfetta. In tale step si deve misurare l'accuratezza delle previsioni fatte negli step precedenti.

Esistono due principali parametri utilizzati per realizzare tale misurazione:

- errore percentuale medio assoluto(MAPE)
- coefficiente di variazione dell'errore(COV)

Il primo si ottiene facendo una media dei valori ottenuti dividendo il valore assoluto dell'errore per la domanda reale e moltiplicando per 100 il risultato. Il secondo si trova dividendo la deviazione standard dell'errore per la domanda reale media. In figura 8[15] il valore del MAPE è 18.31 corrispondente alla media dei valori presenti nella quinta colonna, il valore del COV invece è 0.216.

Generalmente, gli errori di previsione vengono tracciati usando lo stesso intervallo e gli stessi periodi di pianificazione adoperati per i labor driver. In figura 9[15] viene mostrato il plot della variazione di un labor driver e del corrispondente errore di previsione, esso mette in risalto come in periodi caratterizzati da un'alta domanda le previsioni risultino più accurate che nei periodi di bassa domanda.

Exhibit 10					
Tracking forecast error					
	Forecast	Actual	Error	Percentage Error	Absolute Value Percentage Error
Week 1	10	12	2	-16.67	16.67
Week 2	11	8	3	37.50	37.50
Week 3	9	10	1	-10.00	10.00
Week 4	10	11	1	-9.09	9.09
Average		10.25			18.31

Figura 2.8: Calcolo del MAPE

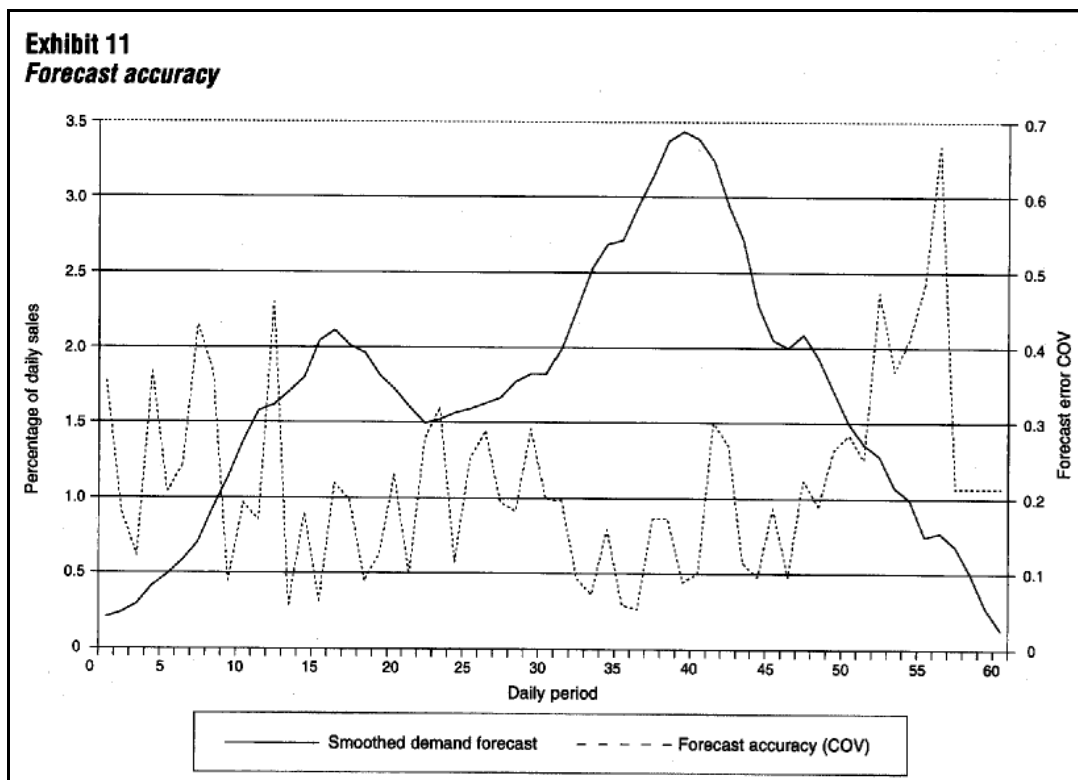


Figura2.9: Plot di un labor driver e del corrispondente errore di previsione

2.2.8. Passo8: definire una finestra di esecuzione per un lavoro controllabile

I driver che determinano il lavoro controllabile, e che quindi devono essere presi in considerazione per determinare la finestra di esecuzione dell'attività, possono essere variabili o meno nel tempo. Naturalmente per determinare l'effettiva finestra di esecuzione si dovrà tener conto sia dei labor driver e per quelli tempo varianti si dovranno realizzare le stesse fasi precedentemente illustrate negli step da 4 a 7 [15].

2.3. Seconda fase: traduzione della richiesta del servizio nella domanda di impiegati

La seconda fase della procedura di scheduling usa come input la previsione della richiesta del servizio determinata nella prima fase e calcola la domanda di impiegati necessaria a soddisfare adeguatamente il cliente nel corso di un prefissato periodo di tempo.

L'output di tale fase, ossia il numero ideale di staff in servizio, costituirà inoltre l'input utilizzato dalla terza fase.

Il risultato del processo di traduzione varia in relazione alla tipologia del lavoro considerato: per un lavoro non controllabile esso consiste nella determinazione del numero ideale di impiegati in servizio in ogni periodo di pianificazione dell'intervallo di riferimento, per un lavoro controllabile esso specifica il carico di lavoro totale e la finestra durante la quale è possibile svolgere l'attività.

Per realizzare la traduzione, generalmente, possono essere usati tre approcci differenti:

- 1) approccio basato sugli standard di produttività**
- 2) approccio basato sugli standard di servizio**
- 3) approccio basato sugli standard economici.**

Il primo approccio costituisce il più facile mezzo per tradurre la richiesta del servizio prevista in fabbisogno di personale. Usando uno standard di produttività si assegna ad ogni impiegato una produttività costante.

Tali standard risultano di facile applicazione per il lavoro controllabile in cui, non essendo obbligatoria l'interazione tra operatore e cliente, è possibile un'esecuzione ininterrotta del lavoro.

Nelle attività caratterizzate da una certa interazione staff-cliente, devono essere necessariamente contemplati i periodi di inattività. Il periodo di inattività pianificato è dato dalla differenza tra il livello massimo di attività e la produttività standard.

Il problema principale connesso all'uso di uno standard di produttività, nel caso di un lavoro non controllabile, consiste nel non poter più considerare costante il livello di produttività di un impiegato. Per ovviare a tale problema si potrebbe alternativamente aumentare in numero di staff nei periodi con maggiore affluenza di clienti mantenendo costante lo standard fissato.

Usando uno standard di servizio si stabilisce un livello di servizio costante in ogni periodo del giorno. Esistono varie opportunità per misurare il livello di servizio, tra esse:

- tempo medio di attesa dei clienti per il servizio
- percentuale di clienti che hanno dovuto aspettare un periodo maggiore rispetto quello standard stabilito per ricevere il servizio.

Nell'uso di tale standard è implicito che il tempo di inattività per impiegato è minore quando il carico di lavoro è elevato piuttosto che quando è basso.

La principale difficoltà dello sviluppo di un simile standard consiste nella determinazione di un appropriato livello di servizio determinato attraverso studi riguardanti le preferenze dei clienti e le caratteristiche del servizio, osservazione diretta e sperimentazioni[17].

Lo scopo di utilizzare uno standard economico è quello di fornire un servizio al minor costo possibile. Ciò significa che si cerca di fornire un servizio migliore nei periodi in cui la domanda per il servizio è alta rispetto a periodi caratterizzati da una bassa domanda. Un servizio migliore è economicamente giustificato nei periodi di alta domanda poichè è provato da più clienti. Al contrario, il costo della distribuzione di un elevato servizio in periodi di bassa domanda potrebbe avere un peso maggiore rispetto ai vantaggi da esso ricavabili.

Gli standard economici hanno diversa complessità. Essi possono essere usati per stimare il costo di attesa dei clienti per il servizio oppure per determinare il livello di staff risultante dalla distribuzione del servizio al più basso costo [18]. Il miglioramento del livello di servizio suggerisce che solitamente gli standard di servizio vengono fissati a un livello troppo basso in molte delle attività che li utilizzano.

2.3.1. Confronto tra i tre approcci usati per la traduzione

Dopo aver introdotto i tipi di standard utilizzati nella traduzione si può eseguire un confronto tra questi per poter determinare quale risulta il più opportuno da utilizzare nel settore di business considerato. A tale scopo sono proposti due ipotetici scenari in base ai quali viene realizzato il confronto.

Scenario 1

Il primo scenario è caratterizzato dalle seguenti condizioni:

- il labor driver è costituito dal numero di clienti
- il salario orario per ogni dipendente è \$10
- si lavora al 100% della capacità
- gli arrivi di clienti in un'ora variano da 0 a 700
- la produttività max per impiegato è fissata a 16 clienti all'ora
- la produttività standard è fissata a 14 clienti all'ora

Per prima cosa si determina il numero di impiegati da schedulare con lo standard di produttività approssimando le frazioni all'intero successivo (ciò significa che se in un'ora arrivano 29 clienti, allora saranno schedati 3 camerieri), in questo modo è possibile determinare quanti impiegati devono essere schedati per ogni volume di clienti in arrivo.

Il livello di staff così determinato genera un tempo medio di attesa di 0.8 minuti e un costo di attesa pari a \$13.46.

Successivamente viene preso in considerazione uno standard di servizio tale che il corrispondente tempo medio di attesa dei clienti non risulti maggiore di quello trovato con lo standard di produttività (ossia 0.8 minuti) e uno standard economico pari al costo di attesa del cliente derivante dall'applicazione dello standard di produttività (ossia \$13.46).

In tal modo sia lo standard di produttività che quello di servizio sono resi concordi allo standard economico.

In figura 10[19] è illustrato il risultato della comparazione tra il livello di staff determinato con l'approccio dello standard economico e i livelli individuati dai rimanenti due approcci. Da essa si evince che:

- il livello di staff specificato dallo standard di produttività tende all'understaffing per molti tassi di arrivo dei clienti
- l'approccio dello standard di servizio tende all'overstaffing per bassi volumi di attività e, in modo crescente, all'understaffing appena i volumi di attività crescono
- il livello di staff specificato dallo standard economico invece si mantiene costante per ogni volume di attività.

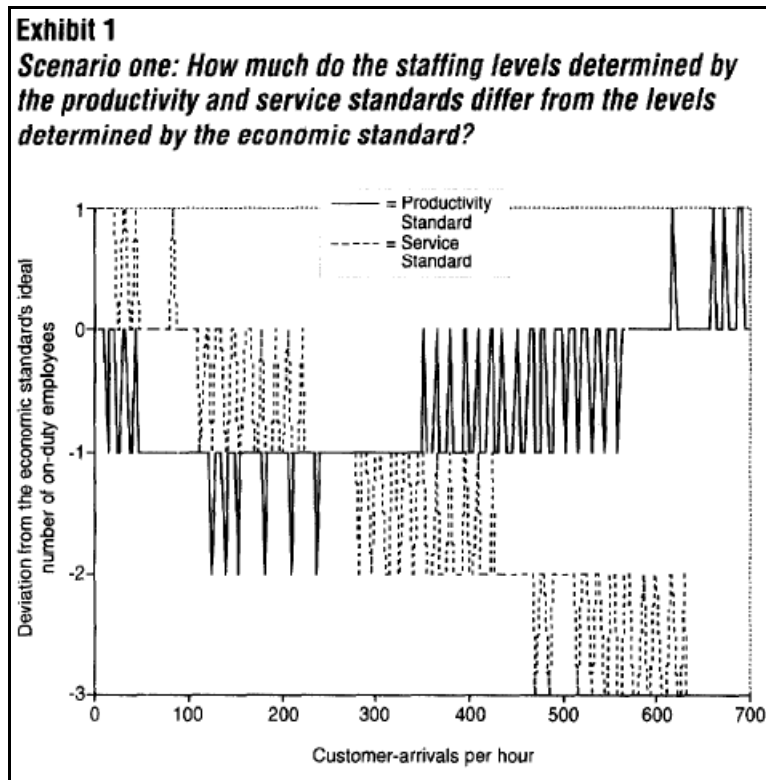
Poiché il livello di staff determinato con il terzo approccio fornisce il migliore risultato economico (si distribuisce il servizio al minor costo possibile), esso rappresenterà il numero ideale di staff in servizio e ogni deviazione da tale livello evidenzierà un risultato economico inferiore. Schedulando un numero inferiore di impiegati, rispetto a quello ideale, i costi del personale si mantengono bassi, tuttavia tali risparmi sono controbilanciati dall'elevato costo di attesa del cliente. Al contrario la schedulazione di un numero di impiegati maggiore rispetto quello ideale comporta un aumento dei costi del personale non giustificato dai benefici ottenuti dalla corrispondente riduzione del tempo di attesa del cliente.

La figura numero 11[19] confronta i risultati economici conseguibili con i tre approcci evidenziando che:

- in un approccio basato sullo standard di servizio sia l'understaffing, corrispondente a un alto tasso di arrivo di clienti, che l'overstaffing, relativo a un basso volume di attività, provocano un abbassamento del risultato economico.
- uno standard di produttività genera costi maggiori, rispetto ad uno standard economico, in un ampio intervallo di tassi di arrivo a causa di un insufficiente servizio associato con il suo frequente understaffing.

- il risultato economico conseguito con gli standard di servizio e di produttività può essere paragonato a quello realizzato mediante uno standard economico solo in un ristretto intervallo di tassi di arrivo.

Ciò implica una maggiore utilità dello standard economico rispetto agli altri due.



**Figura 2.10: Confronto tra i livelli di staff determinati con i tre approcci
nel primo scenario**

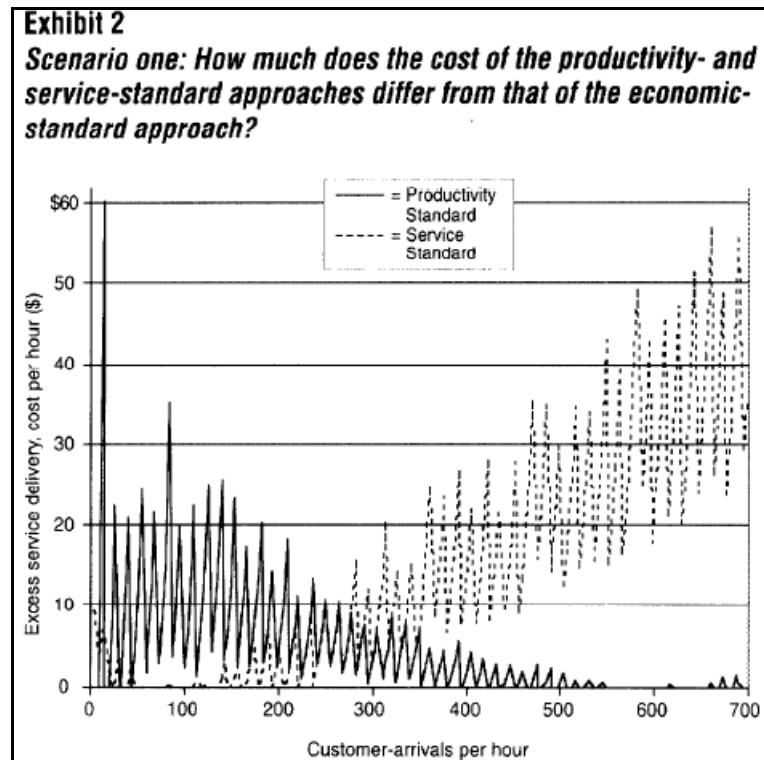


Figura 2.11: Confronto tra i costi generati dai tre approcci nel primo scenario di riferimento

Scenario 2

Il secondo scenario considerato è caratterizzato dalle seguenti condizioni:

- il labor driver è costituito dal numero di clienti
- il salario orario per ogni dipendente è \$10
- si lavora al 100% della capacità
- gli arrivi di clienti in un'ora variano da 0 a 700
- i valori degli standard di produttività e di servizio sono raddoppiati rispetto al caso precedente
- il costo associato all'attesa del cliente è di \$26.92, quindi doppio rispetto al caso precedente

Considerando che questa volta gli standard di produttività e di servizio sono in contrasto con il valore economico dell'attesa del cliente si procede al confronto tra i livelli di staff determinati con i tre approcci e rappresentati in figura 12[19].

Appare evidente che:

- gli standard di produttività e quello di servizio generano un sottodimensionamento del personale se paragonati allo standard economico (rappresentativo del numero ideale di staff da schedare)
- fissando un livello di servizio più alto si crea un fenomeno di overstaffing per bassi tassi di arrivo dei clienti e un aumento dell'understaffing (fino 5 unità in meno) man mano che cresce il volume di attività.

Ancora una volta lo standard economico si mostra più efficace rispetto i rimanenti, tuttavia nella pratica esso non è usato così ampiamente come dovrebbe, probabilmente a causa della semplicità di applicazione dello standard di produttività.

I livelli di staff basati sugli standard di produttività possono essere calcolati a mente, oppure su un singolo foglio di carta. Al contrario gli standard economici richiedono per il loro uso modelli di coda o di simulazione.

Paragonando, ancora una volta, i costi prodotti dai tre approcci attraverso un diagramma riportato in figura 13[19] si nota che:

- sia gli standard di produttività che di servizio sono consistentemente più costosi rispetto quelli economici
- solo in un ristretto intervallo di tassi di arrivo dei clienti i costi generati dall'uso di standard di produttività o di servizio possono eguagliare quelli relativi all'adozione di uno standard economico
- i costi in eccesso legati agli standard di produttività e di servizio sono persino più alti che nel primo scenario

Concludendo è possibile affermare che:

- Gli standard di produttività e di servizio eguagliano quelli economici solo per un ristretto intervallo di volume di attività. Unicamente in tale intervallo è possibile realizzare una corretta applicazione di questi due approcci, il principale ostacolo per tale implementazione è costituito proprio dall'identificazione dell'intervallo in questione.
- Gli standard di produttività e di servizio sono molto inaccurati specialmente quando corrispondono a particolari costi di attesa per il cliente. Tale inaccuratezza porta a livelli di staff maggiori o minori rispetto quelli ideali e, di conseguenza, a costi di distribuzione più alti rispetto quelli derivanti dall'uso di standard economici.
- Solo uno standard economico può mettere in luce gli effetti economici legati alla realizzazione di buon servizio o di un cattivo servizio.

Per tali ragioni, a dispetto della difficoltà connessa al loro uso, lo standard economico risulta il più appropriato nel settore applicativo considerato.

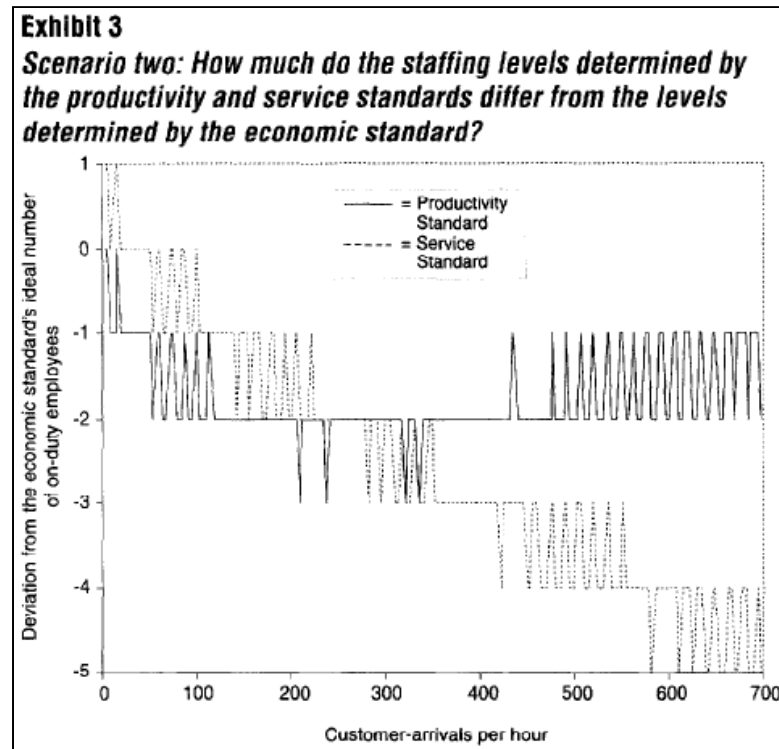


Figura 2.12: Confronto tra i livelli di staff determinati con i tre approcci nel secondo scenario

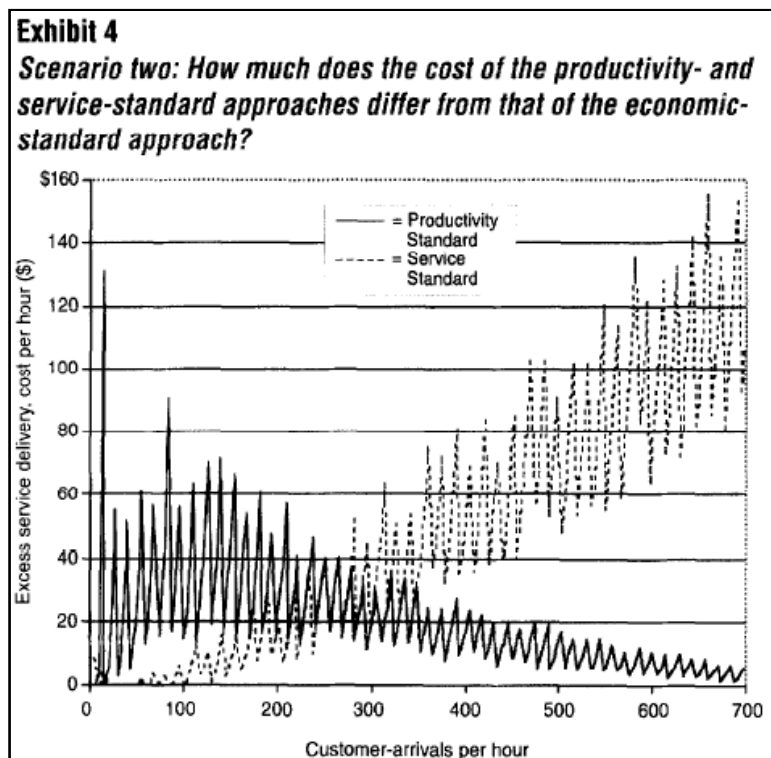


Figura 2.13: Confronto tra i costi generati dai tre approcci nel secondo scenario di riferimento

2.3.2. Sviluppo di uno standard economico

Per lo sviluppo di un efficiente standard economico esistono varie alternative. Di seguito ne saranno illustrate due considerando il seguente scenario di riferimento:

- Ogni cameriere può servire 16 clienti in un'ora
- In una data ora ci si aspetta l'arrivo di 112 clienti
- Il costo di un'ora di lavoro per ogni impiegato è di \$10
- Il costo di una transazione è di \$5

Metodo I: uso del costo di attesa del cliente

Una possibile alternativa per sviluppare uno standard economico è quella di stimare un costo per l'attesa del cliente. Poichè, dalle ipotesi fatte, risulta che il costo di una transazione è inferiore al costo di un'ora di

lavoro, è possibile assumere per il costo orario di attesa del cliente un valore pari a \$10 (ossia al costo di un'ora di lavoro per impiegato).

Applicando le ipotesi alla base della teoria delle code rappresentate in figura 15[19] è stata realizzata la tabella in figura 14[19]. Essa fornisce il numero ideale di impiegati da schedulare corrispondente al minimo costo per la distribuzione del servizio (pari a 9 in tal caso) ed è costituita dalle seguenti colonne:

- Colonna n°1: numero ideale di impiegati
- Colonna n°2: tempo medio di attesa per il singolo cliente espresso in minuti
- Colonna n°3: tempo totale di attesa espresso in ore e ottenuto moltiplicando il tempo medio di attesa per il singolo cliente per il numero totale di clienti arrivati
- Colonna n°4: costo totale di attesa espresso in dollari (\$), ottenuto moltiplicando il tempo totale di attesa per il costo orario di attesa fissato
- Colonna n°5: costo del lavoro espresso in dollari, ottenuto moltiplicando il numero ideale di impiegati per il costo di un'ora di lavoro
- Colonna n°6: costo totale di distribuzione del servizio espresso in dollari, ottenuto sommando il costo del lavoro e il costo totale di attesa.

Uno dei problemi connessi all'uso di questa tecnica è quello di considerare una relazione lineare tra il tempo e il costo di attesa. Ciò significa che il costo dell'attesa di 100 clienti per un tempo pari a un minuto risulta identico al costo dell'attesa di un cliente per un tempo di 100 minuti.

Exhibit 5 <i>Applying an economic standard</i>					
Number of Servers	Wq^a	Total waiting time^b	Total waiting cost^c	Labor cost^d	Total cost^e
8	2.382	4.447	\$44.47	\$80.00	\$124.47
9	0.722	1.347	\$13.47	\$90.00	\$103.47
10	0.277	0.517	\$5.17	\$100.00	\$105.17

Figura 2.14: Applicazione di uno standard economico basato sulla teoria delle code

Using a Queuing Model

There are two key tools for determining employee requirements with service standards and economic standards: queuing models and simulation.

Queuing models are formulas that describe the performance of a service system. Many queuing models have been developed using various sets of assumptions. One of the most famous queuing models describes a multiple-server, single-queue system. The model has three inputs: the customer-arrival rate (λ); the rate at which a server can serve customers (μ); and the number of servers (m). These formulas describe the performance of the system:

$$\text{Server utilization: } \rho = \frac{\lambda}{m\mu}$$

$$\text{Probability of zero customers in the system: } P_0 = \left[\sum_{n=0}^{m-1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^m}{m!} \frac{1}{1-\rho} \right]^{-1}$$

$$\text{Probability of } n \text{ customers in the system: } P_n = \begin{cases} \frac{(\lambda/\mu)^n}{n!} P_0, & \text{if } 0 \leq n \leq m \\ \frac{(\lambda/\mu)^n}{m! m^{n-m}} P_0, & \text{if } n \geq m \end{cases}$$

$$\text{Average number of customers in the queue: } L_q = \frac{P_0 (\lambda/\mu)^m \rho}{m! (1-\rho)^2}$$

$$\text{Average time a customer spends in the queue: } W_q = \frac{L_q}{\lambda}$$

$$\text{Average number of customers in the system: } L_s = L_q + \frac{\lambda}{\mu}$$

$$\text{Average time a customer spends in the system: } W_s = W_q + \frac{1}{\mu}$$

$$\text{Probability that a customer's wait exceeds } t: P(W_q > t) = (1 - \sum_{n=0}^{m-1} P_n) e^{-m\mu(1-\rho)t}$$

Take, for example, a case in which customers arrive at the rate of 58.7 an hour, each employee can serve 16 customers an hour, and there are five servers:

Server utilization = 0.734 (73.4%)

Probability of zero customers in the system = 0.021 (2.1%)

Average number of customers in the queue = 1.19

Average time a customer spends in the queue = 0.020 hrs (1.22 min)

Average number of customers in the system = 4.86

Average time a customer spends in the system = 0.083 hrs (4.97 min)

$P(\text{wait time exceeds three minutes [0.05 hours]}) = 0.149$

The second key tool for determining employee requirements with service standards and economic standards is simulation. Although it is more robust, it is harder to implement.—G.M.T.

Source: F.S. Hillier and G.S. Lieberman, *Introduction to Operations Research* (New York: McGraw Hill, 1995), pp. 683–684.

Figura 2.15: Modello della teoria delle code

Metodo II: uso di una focalizzazione sulle rendite

Tale approccio permette di superare i limiti imposti dal precedente. In esso risulta fondamentale individuare il tempo di attesa per il quale si verifica la perdita del cliente. Ciò può essere fatto attraverso l'osservazione, la sperimentazione, l'esperienza. Si prenda in considerazione, per il momento, soltanto la perdita di vendite correnti e si supponga di aver osservato che per un tempo di attesa inferiore o uguale a 10 minuti non si registrano perdite di clienti, riscontrabili solo per un tempo di attesa superiore a tale limite. In questo caso il numero ideale di staff corrisponde a 9 unità come segnalato in figura 16[19]. Al valore 9 corrisponde infatti il massimo profitto netto(\$468.96).

La tabella in questione è caratterizzata dalle seguenti colonne:

- Colonna n°1: numero ideale di impiegati
- Colonna n°2: probabilità che un cliente debba attendere per un periodo minore o uguale a 10 minuti
- Colonna n°3: probabilità che un cliente debba attendere per un periodo superiore a 10 minuti
- Colonna n°4: transazioni effettuate ottenute moltiplicando la probabilità espressa nella seconda colonna per il numero totale di clienti (112)
- Colonna n°5: transazioni perse ottenute moltiplicando la probabilità espressa nella terza colonna per il numero totale di clienti (112)
- Colonna n°6: valore totale delle transazioni espresso in dollari, ottenuto moltiplicando la differenza tra le transazioni effettuate e quelle perse per il valore di ogni transazione (\$5)
- Colonna n°7: costo del lavoro espresso in dollari, ottenuto moltiplicando il numero ideale di impiegati per il costo di un'ora di lavoro
- Colonna n°8: profitto netto espresso in dollari, ottenuto dalla differenza tra il valore totale delle transazioni e il costo del lavoro.

Si prenda in considerazione, adesso, una transazione di \$100. In tal caso il numero di camerieri che massimizza il profitto netto diventa pari a 10 come mostrato in figura 17[19]. Generalmente, maggiore è il costo della transazione, più alto è il numero di camerieri economicamente vantaggioso.

L'approccio illustrato può essere ulteriormente raffinato identificando, piuttosto che un singolo punto di taglio oltre il quale si verifica la perdita totale delle vendite, la quota di transazioni perse all'interno di una serie di intervalli di attesa contenuti nell'intervallo da 0 a 10 minuti precedentemente considerato. In figura 18[19] sono considerati i seguenti intervalli:

- Da 0 a 3 minuti: in esso si registra una perdita di vendite pari allo 0%
- Da 3 a 5 minuti: in esso si registra una perdita di vendite pari al 20%
- Da 5 a 10 minuti: in esso si registra una perdita di vendite pari al 60%
- Superiore a 10 minuti: in esso si registra una perdita di vendite pari al 100%.

Il calcolo delle transazioni perse viene effettuato sommando le transazioni perse in ogni intervallo di tempo considerato e moltiplicando tale valore per il numero di clienti attesi, ossia 112.

Le transazioni perse in un singolo intervallo sono date dal prodotto tra la probabilità di un tempo di attesa che ricade nell'intervallo considerato per la corrispondente percentuale di vendite perdute.

Considerando il caso di 8 camerieri in servizio il calcolo delle transazioni perse si realizza con la seguente formula:

$$[(0.715 \times 0.0) + (0.118 \times 0.2) + (0.123 \times 0.6) + (0.044 \times 1.0)] \times 112 = 15.875$$

Il calcolo delle transazioni effettuate è dato da:

$$\{1 - [(0.715 \times 0.0) + (0.118 \times 0.2) + (0.123 \times 0.6) + (0.044 \times 1.0)]\} \times 112 = 96.125$$

Exhibit 6 <i>Applying a revenue focus</i>							
Number of servers	$P(W \leq 10)^a$	$P(W > 10)^b$	Transactions made ^c	Transactions lost ^d	Total transaction value ^e	Labor cost ^f	Net benefit ^g
8	0.9559	0.0441	107.056	4.944	\$535.28	\$80.00	\$455.28
9	0.9981	0.0019	111.792	0.208	\$558.96	\$90.00	\$468.96
10	0.9999	0.0001	111.992	0.008	\$559.96	\$100.00	\$459.96

Figura 2.16: Applicazione di uno standard economico basato su una focalizzazione sulle rendite

Exhibit 7 <i>Transaction-contribution effect</i>							
Number of servers	$P(W \leq 10)^a$	$P(W > 10)^b$	Transactions made ^c	Transactions lost ^d	Total transaction value ^e	Labor cost ^f	Net benefit ^g
8	0.9559	0.0441	107.056	4.944	\$10,705.59	\$80.00	\$10,625.59
9	0.9981	0.0019	111.792	0.208	\$11,179.18	\$90.00	\$11,089.18
10	0.9999	0.0001	111.992	0.008	\$11,199.17	\$100.00	\$11,099.17
11	1.0000	0.0000	112.000	0.000	\$11,200.00	\$110.00	\$11,090.00

Figura 2.17: Effetti di un aumento del costo di una transazione

Exhibit 8**Accounting for different customer-waiting times**

	Number of servers			
	8	9	10	11
(1) $P(W \leq 3)^a$	0.7146	0.9222	0.9798	0.9951
(2) $P(3 < W \leq 5)^b$	0.1180	0.0510	0.0161	0.0043
(3) $P(5 < W \leq 10)^c$	0.1233	0.0249	0.0040	0.0006
(4) $P(W > 10)^d$	0.0441	0.0019	0.0001	0.0000
Transactions lost ^e	15.875	3.022	0.636	0.137
Transactions made ^f	96.125	108.978	111.364	111.863
Total transaction value ^g	\$480.63	\$544.89	\$556.82	\$559.32
Labor cost ^h	\$80.00	\$90.00	\$100.00	\$110.00
Net benefit ⁱ	\$400.63	\$454.89	\$456.82	\$449.32

Figura 2.18: Identificazione di più intervalli di valutazione per stimare le vendite perdute

Perdita di vendite future

Partendo dalla considerazione che un cattivo servizio non solo incide sulle transazioni correnti, ma influenza negativamente anche quelle successive, e che un servizio veramente buono può incrementare consistentemente il volume futuro di business, è possibile realizzare un'analisi ancora più accurata prendendo in considerazione l'effetto sulle vendite future determinato dalle caratteristiche del servizio.

Si suppone che, in base alla diretta osservazione e all'esperienza, la relazione esistente tra il tempo di attesa e le transazioni future sia rappresentabile attraverso la tabella 1[19]. È possibile fare delle considerazioni sui casi limite presenti in tabella:

- Se il cliente viene servito in meno di 0.15 minuti, egli percepisce il servizio come eccellente, ciò provoca un aumento di 0.5 transazioni, in media, per ogni cliente
- Un servizio veramente scadente (con un tempo di attesa superiore a 10 minuti) ha come risultato la perdita di ben 2 transazioni per ogni cliente. Esse si realizzano attraverso un mancato ritorno del cliente o attraverso la diffusione di negativi giudizi verbali che possono influenzare altri potenziali clienti.

Alla luce di tali considerazioni si è effettuata una nuova stima per determinare il giusto numero di personale in servizio, riportata in figura 19[19]. Il livello di impiegati che procura il massimo profitto netto è 12. Il calcolo delle transazioni perse e di quelle realizzate si effettua in modo analogo al caso precedente. Ad esempio, considerando una schedulazione di 8 impiegati il valore delle transazioni perse è dato da:

$$[(0.325 \times 0.0) + (0.118 \times 0.2) + (0.123 \times 0.6) + (0.044 \times 2.0)] \times 112 = 20.819$$

Quello delle transazioni realizzate si ricava dalla seguente espressione:

$$\{1 - [(0.325 \times 0.0) + (0.118 \times 0.2) + (0.123 \times 0.6) + (0.044 \times 2.0)] + (0.340 \times 0.5)\} \times 112 = 21.817$$

Solitamente, maggiore è l'orizzonte temporale influenzato dalle caratteristiche del servizio, maggiore è il livello di staff da utilizzare.

Poichè una stima di tali effetti risulta molto complicata i manager sono spesso portati a sottovalutarne l'importanza, determinando livelli di staff più bassi rispetto quelli che effettivamente massimizzano il profitto.

Tabella 2.1: Relazione tra tempo di attesa e futuro business

Wait time in minutes (W)	Effect on future transactions
$W \leq 0.15$	0.5
$0.15 < W \leq 3$	0.0
$3 < W \leq 5$	-0.2
$5 < W \leq 10$	-0.6
$10 < W$	-2.0

Exhibit 9						
Accounting for waiting effects on future business						
	Number of servers					
	8	9	10	11	12	13
(1) $P(W \leq 0.15)^a$	0.3896	0.6446	0.8033	0.8968	0.9713	0.9759
(2) $P(0.15 < W \leq 3)^b$	0.3250	0.2776	0.1765	0.0983	0.0284	0.0239
(3) $P(3 < W \leq 5)^c$	0.1180	0.0510	0.0161	0.0043	0.0003	0.0002
(4) $P(5 < W \leq 10)^d$	0.1233	0.0249	0.0040	0.0006	0.0000	0.0000
(5) $P(10 < W)^e$	0.0441	0.0019	0.0001	0.0000	0.0000	0.0000
Transactions lost ^f	20.819	3.231	0.644	0.137	0.029	0.006
Transactions gained ^g	21.817	36.100	44.987	50.221	53.129	54.651
Net transactions ^h	112.998	144.870	156.343	162.084	165.010	166.645
Total transaction value ⁱ	\$564.99	\$724.35	\$781.71	\$810.42	\$825.50	\$833.22
Labor cost ^j	\$80.00	\$90.00	\$100.00	\$110.00	\$120.00	\$130.00
Net benefit ^k	\$484.99	\$634.35	\$681.71	\$700.42	\$705.50	\$703.22

Figura 2.19: Determinazione del numero di impiegati da schedulare

2.3.3 Aspetti fondamentali per l'applicazione di uno standard economico

Quando si applica uno standard economico è necessario considerare alcuni fattori talvolta contraddittori.

Errore atteso

Durante la traduzione della richiesta del servizio nel fabbisogno di staff è necessario prendere in considerazione l'errore atteso, poichè all'aumento di questo è legata una maggiore richiesta di personale. Ad esempio, un elevato livello dell'errore può generare una richiesta di staff superiore del 50% rispetto quella effettivamente necessaria. Ciò provoca un aumento dei costi per la distribuzione del servizio di oltre il 39%.

Gli effetti causati da inaccuratezza nelle previsioni evidenziano la stretta connessione tra le quattro fasi dello scheduling della forza lavoro, invitando a realizzare attentamente ogni singola fase in vista della difficoltà di correggere errori generati nelle fasi precedenti.

Slittamento del lavoro

Il processo di traduzione non può essere svolto separatamente per ogni periodo di pianificazione. Ciò appare evidente considerando il seguente esempio relativo ad un ristorante "quick-service" dove:

- Il periodo di pianificazione adottato è di 15 minuti
- Il tempo necessario per la distribuzione del servizio è di 3 minuti.

Si prenda in esame come periodo di pianificazione corrente quello che va dalle 12:45 alle 13:00.

Per un cliente che arriva alle 12:50 la distribuzione del servizio sarà completata durante il periodo corrente, per un cliente che arriva tra le 12:57 e le 13:00 tale distribuzione si estenderà anche a parte del periodo successivo. In tal modo la richiesta del servizio per il periodo successivo subirà un incremento proporzionale al numero di arrivi di clienti nella fascia oraria suddetta. Questo fenomeno, chiamato "slittamento", si ripete per tutti i periodi di pianificazione e può generare la necessità di aumentare il numero di personale in servizio nei periodi successivi quello considerato[20].

Scostamenti dall'ideale

Spesso risulta difficile eguagliare il numero di impiegati in servizio al numero di impiegati effettivamente necessario. Per tale ragione, oltre a determinare il numero ideale di forza lavoro da schedare, è importante valutare i costi derivanti dagli scostamenti da una misura ideale di staff.

Una simile valutazione è realizzata in figura 20[19], caratterizzata dalle seguenti colonne:

- Colonna n°1: periodi di pianificazione
- Colonna n°2: tasso di arrivo dei clienti per ora
- Colonna n°3: numero ideale di staff in servizio per periodo
- Colonna n°4: costo dovuto all'understaffing di 2 impiegati

- Colonna n°5: costo dovuto all'understaffing di 1 impiegato
- Colonna n°6: costo dovuto all'overstaffing di 1 impiegato
- Colonna n°7: costo dovuto all'overstaffing di 2 impiegati.

Per la determinazione di tali dati sono state considerate le condizioni imposte per il primo scenario, ossia:

- il labor driver è costituito dal numero di clienti
- il salario orario per ogni dipendente è \$10
- si lavora al 100% della capacità
- gli arrivi di clienti in un'ora variano da 0 a 700
- la produttività max per impiegato è fissata a 16 clienti l'ora
- il costo di attesa di un cliente è pari a \$13.46.

Le informazioni contenute nella tabella mostrano che:

- Nel caso si voglia aumentare lo staff di un'unità risulta consigliabile farlo nel terzo periodo, poichè ad esso corrisponde un valore più basso del costo di overstaffing
- Nel caso in cui si voglia ridurre lo staff di un'unità bisogna farlo nel secondo periodo
- Non è consigliabile l'understaffing di 2 unità nel primo periodo poichè 3 impiegati potrebbero occuparsi al massimo di 48 clienti e il tasso di arrivo di clienti orario è 50.8.

Per un lavoro controllabile, una volta stabilita l'ideale finestra di esecuzione, i costi dovuti a una deviazione dall'ideale vengono valutati ipotizzando incrementi o decrementi della durata di tale finestra.

Exhibit 11						
<i>Cost of deviating from the ideal staff size</i>						
Period	Customer- arrival rate	Ideal staff size	Number of employees above or below the "ideal staff size"			
			-2	-1	+1	+2
1	50.8	5	NA*	\$13.72	\$5.27	\$13.95
2	74.4	7	\$122.04	\$5.15	\$5.69	\$14.22
3	118.2	10	\$93.62	\$8.69	\$3.58	\$11.00

Figura 2.20: Valutazione dei costi derivanti da uno scostamento dal numero ideale di staff

Feedback e assenteismo

Poichè le quattro fasi dello scheduling sono fortemente collegate tra loro è necessario operare dei feedback durante la loro esecuzione, al fine di poter cogliere aspetti importanti dell'una che possono condizionare il

buon esito dell'altra. Infine non bisogna sottovalutare l'importanza di rintracciare l'assenteismo dei dipendenti durante i giorni della settimana, durante le ore del giorno o in relazione alla specifica mansione, poichè, man mano che esso aumenta, bisogna schedulare un maggior numero di impiegati

2.4. Terza fase: lo scheduling degli impiegati

Scopo finale è sviluppare una schedulazione del lavoro che fornisca un numero sufficiente di personale con le necessarie competenze per realizzare la distribuzione del servizio, considerando, al tempo stesso, vincoli legati alle preferenze degli impiegati, alle normative del governo e a regole specifiche imposte dall'organizzazione.

Le procedure di "workforce scheduling" di solito prendono in considerazione i seguenti obiettivi:

- Minimizzare il costo della schedulazione assicurando un certo livello di staffing
- Massimizzare il beneficio totale ottenibile dalla schedulazione.

Sfortunatamente, una schedulazione che ottimizza obiettivi come la minimizzazione dei costi o la massimizzazione dei benefici, spesso, può scontrarsi con le esigenze dei dipendenti.

La maggior parte delle procedure per la schedulazione della forza lavoro utilizzano uno dei due sistemi classici di base:

- Il modello di Dantzig
- Il modello di Keith.

2.4.1. I sistemi classici di "workforce scheduling"

Modello di Dantzig

Il primo modello preso in considerazione per lo sviluppo di un "workforce scheduling" è quello presentato nel 1954 da George Dantzig. Tale modello, per semplicità indicato come "D-framework", è mostrato in figura 21[14]. In esso:

- T rappresenta il set di turni che possono essere schedulati
- c_t rappresenta il costo dell'assegnazione di un dipendente al turno
- x_t rappresenta il numero di dipendenti che lavorano in un turno t
- a_{tp} rappresenta una variabile di valore 1 se p è un periodo di lavoro per il turno t o di valore 0 in caso contrario

- P rappresenta il set di periodi di pianificazione in un orizzonte di pianificazione settimanale
- r_p rappresenta il numero di dipendenti che l'amministrazione desidera avere in servizio al periodo p
- l'obiettivo (1) è la minimizzazione del costo della schedulazione
- il vincolo (2) assicura che il numero di dipendenti schedulati sia maggiore o uguale al numero dei dipendenti richiesti in ogni periodo, cioè vieta l'understaffing
- il vincolo (3) impone che la variabile decisionale assuma valori interi maggiori o uguali a zero.

Il modello, tramite la sua funzione obiettivo, cerca di eliminare, quando possibile, l'overstaffing. Infine, mediante la condizione di vincolo, proibisce il sottodimensionamento del personale.

Nel D-framework possono essere rilevati due difetti:

- 1) Non viene presa in considerazione, nello sviluppo della schedulazione, la reale disponibilità dei dipendenti durante l'orizzonte di pianificazione. In questo modo, è possibile che la schedulazione realizzata risulti inutilizzabile a causa della non disponibilità di qualche impiegato.
- 2) La proibizione dell'understaffing. Molti manager che utilizzano il modello in questione spesso evitano di applicare il vincolo che proibisce il sottodimensionamento del personale nei periodi in cui non c'è l'overstaffing per poter ottenere un miglior bilanciamento dei costi.

$\text{Min } Z = \sum_{t \in T} c_t x_t, \quad (1)$ <p style="text-align: center;">subject to</p> $\sum_{t \in T} a_{tp} x_t \geq r_p \quad \text{for } p \in P, \quad (2)$ $x_t \geq 0 \text{ and integer for } t \in T. \quad (3)$
--

Figura 2.21: Modello di Dantzig

Modello di Keith

Nato per rimediare ad alcuni degli inconvenienti presenti nel modello precedente, il cosiddetto “K-framework”, proposto da Elbridge Keith nel 1979, è rappresentato in figura 22[14]. In esso:

- α_p è una variabile che rappresenta un understaffing incontrollabile nel periodo p
- β_p è una variabile che rappresenta un understaffing controllabile nel periodo p
- σ_p è una variabile che rappresenta un overstaffing controllabile nel periodo p

- π_p è una variabile che rappresenta un overstaffing incontrollabile nel periodo p
- b_p^β è una variabile che rappresenta un limite per l'understaffing nel periodo p
- b_p^σ è una variabile che rappresenta un limite per l'overstaffing nel periodo p
- l'obiettivo (4) è quello di minimizzare il costo della schedulazione
- il vincolo (5) assicura che il numero di dipendenti in servizio sia uguale a quello dei dipendenti realmente necessari effettuando una correzione sulle variabili che rappresentano il sovradimensionamento o il sottodimensionamento del personale
- i vincoli (6) e (7) impongono limiti superiori alle variabili che rappresentano l'understaffing e l'overstaffing
- il vincolo (8) assicura che per ogni periodo lavori almeno un impiegato
- i vincoli (9) (10) impongono che le variabili decisionali siano intere e maggiori o uguali a zero.

Questo metodo realizza una schedulazione cercando di minimizzare i costi reali di distribuzione del servizio e gli “pseudo-costi” (penalità) associati all'understaffing e all'overstaffing.

Il vincolo (8) non è originariamente presente nel modello. Tale condizione viene aggiunta per impedire l'estremo understaffing (ossia una situazione in cui nessun dipendente viene schedulato in un dato periodo). In questo modo il modello di Keith si mostra più adatto per risolvere i problemi di schedulazione rispetto al modello di Dantzig precedentemente analizzato.

$$\begin{aligned}
 \text{Min } Z &= \sum_{t \in T} c_t x_t \\
 &+ \sum_{p \in P} [10\alpha_p + 3\beta_p + 0\sigma_p + 0.5\pi_p], \quad (4) \\
 \text{subject to} \\
 \sum_{t \in T} a_{tp} x_t + \alpha_p + \beta_p - \sigma_p - \pi_p &= r_p \quad \text{for } p \in P, \quad (5) \\
 \beta_p &\leq b_p^\beta \quad \text{for } p \in P, \quad (6) \\
 \sigma_p &\leq b_p^\sigma \quad \text{for } p \in P, \quad (7) \\
 \alpha_p &\leq r_p - b_p^\beta - 1 \quad \text{for } p \in P, \quad (8) \\
 \alpha_p, \beta_p, \sigma_p, \pi_p &\geq 0 \text{ and integer} \quad \text{for } p \in P, \quad (9) \\
 x_t &\geq 0 \text{ and integer} \quad \text{for } t \in T. \quad (10)
 \end{aligned}$$

Figura 2.22: Modello di Keith

Limiti dei modelli classici

Malgrado la generale superiorità del K-framework rispetto al D-framework, entrambi i modelli presentano delle limitazioni che riducono la loro efficacia:

- 1) **Paradigma del singolo periodo:** i sistemi basati su tale paradigma usano le richieste di personale in modo indipendente per ogni singolo periodo di pianificazione. In questo modo vengono generati dei turni discontinui, ossia turni che interessano periodi di pianificazione non consecutivi, che cominciano e finiscono in particolari orari. Ciò costituisce un problema poichè solitamente i turni del personale devono interessare periodi di pianificazione successivi. Un simile approccio impedisce che una schedulazione distribuisca, in un determinato periodo di pianificazione, un numero di impiegati minore di quello ideale (ma accettabile per il livello di servizio) al fine di ottenere un maggiore contenimento dei costi e una migliore schedulazione in altri periodi.
- 2) **Credenza che il valore di un dipendente extra sia indipendente dal momento in cui egli viene inserito:** l'aggiunta di un dipendente è strettamente legata al periodo in cui viene realizzata. Aggiungere un dipendente extra in un periodo di piena attività potrebbe produrre sostanziali miglioramenti al servizio, mentre inserirlo in un periodo poco produttivo comporterebbe vantaggi minimi.

2.4.2. I nuovi modelli

Per rimediare ai difetti dei modelli classici sono stati ideati due nuovi sistemi di seguito citati:

- Contemporaneo modello di servizio o “CS-framework”
- Contemporaneo modello economico o “CE-framework”.

Tali modelli superano le limitazioni dei modelli classici rispettivamente nell'ambito dello standard di servizio e in quello dello standard economico. Ad esempio, l'obiettivo del “CE-framework” è quello di realizzare una schedulazione che massimizza il profitto economico totale procedendo prima con la schedulazione del numero minimo accettabile di dipendenti in ogni periodo di pianificazione, e determinando poi il numero aggiuntivo di dipendenti che potrebbe fornire un incremento nella performance economica. Questo modello cerca di trovare una soluzione che bilancia al meglio il beneficio monetario derivante da un buon servizio con il costo monetario legato ad un servizio inadeguato e alla relativa distribuzione. Bisogna notare che solitamente il numero minimo di dipendenti è inferiore a quello ideale. Il “CE-framework” opera in modo analogo nel campo dello standard economico.

Le principali differenze tra tali modelli e i precedenti riguardano la seconda e la terza fase della procedura di scheduling, ossia la determinazione del fabbisogno ideale di staff e la creazione della schedulazione. I modelli classici prima determinano il numero ideale di staff e poi lo usano come input per compilare la schedulazione tenendo conto delle preferenze dei dipendenti in modo indipendente per ogni singolo

periodo di pianificazione. I nuovi modelli, invece, considerano le preferenze del personale contemporaneamente all'esecuzione delle suddette fasi, riconoscendo l'impossibilità di realizzare un'efficiente schedulazione sottovalutando la forte interdipendenza tra le esigenze dello staff e i diversi periodi di pianificazione. Il "CE-framework" e il "CS-framework", caratterizzati da un paradigma attraverso i periodi ("cross-period paradigm"), producono migliori schedulazioni rispetto ai modelli classici poiché fanno un uso più appropriato delle informazioni disponibili.

2.4.3. Sviluppo dello scheduling della forza lavoro

Nello sviluppo di una schedulazione un manager può utilizzare:

- I. una procedura a due fasi;**
- II. una procedura a una fase.**

Le procedure a due fasi impongono l'uso del paradigma classico del singolo periodo. Esse sono così strutturate: nella prima fase viene sviluppata una schedulazione senza prendere in considerazione i desideri o le disponibilità dei dipendenti; nella seconda fase, sulla base della schedulazione effettuata si assegnano i turni ai dipendenti, rispettando, se è possibile, le loro preferenze. Gli approcci a due fasi sono molto veloci e convenienti da sviluppare, ma non sempre sono in grado di soddisfare alcune richieste della schedulazione, come quella di assicurare che un dipendente lavori un minimo numero di ore.

Le procedure a una fase, invece, tengono conto delle informazioni sulle preferenze del personale durante il processo di sviluppo della schedulazione. Tali tipi di approcci, a meno che non siano ben strutturati, possono risultare molto lenti, tuttavia il loro uso diventa sempre più comune a causa dei problemi connessi alle procedure a due fasi.

Confronto tra i due tipi di procedura

Per poter cogliere le principali differenze tra i due approcci si analizzeranno le soluzioni da essi prodotte per la risoluzione di un semplice problema di scheduling. Si assume che:

- l'orizzonte temporale di scheduling è di 8 ore
- i periodi di pianificazione sono 8 da un'ora ciascuno
- la durata di un turno varia da 4 a 8 ore
- i dipendenti sono 2: A e B
- A è disponibile dall'ora 1 all'ora 7
- B è disponibile dall'ora 3 all'ora 8

- L'azienda necessita di un impiegato durante le prime e le ultime due ore, e di entrambi durante le ore centrali.

Queste informazioni sono schematizzate nella parte superiore delle due tabelle in figura 23[21]. Nella parte inferiore delle tabelle sono rappresentate, invece, le due soluzioni al problema.

La soluzione ottenuta con l'approccio a due fasi schedula il dipendente A per 8 ore e il dipendente B per le 4 ore centrali. Questa schedulazione incontra gli standard del livello di servizio bilanciando il numero di dipendenti necessari con quello di dipendenti schedulati, come indicato dagli 0 nella riga "net staffing level", tuttavia non può essere attuata poichè non riesce soddisfare le esigenze dei dipendenti infatti: A viene schedulato in un periodo di tempo in cui non può lavorare, il periodo 8, e B otterrà uno stipendio poco soddisfacente, corrispondente alle 4 ore lavorate. La schedulazione ottenuta con l'approccio a una fase schedula l'impiegato A per sei ore, dal periodo 1 al 6 e l'impiegato B per sei ore che vanno dal periodo 3 al periodo 8. Essa dunque terrà conto delle disponibilità dei dipendenti, bilanciando gli stipendi per entrambi e incontrando, al tempo stesso, gli standard di servizio.

Two-phase (shift-based) solution								
Hour	1	2	3	4	5	6	7	8
Employees needed	1	1	2	2	2	2	1	1
Employee A availability	Y	Y	Y	Y	Y	Y	Y	N
Employee B availability	N	N	Y	Y	Y	Y	Y	Y
Shift-based schedule (two-phase)	Shift 1 (employee A)	W	W	W	W	W	W	W
	Shift 2 (employee B)			W	W	W		
	Total scheduled	1	1	2	2	2	2	1
	Net staffing level	0	0	0	0	0	0	0
One-phase (holistic) solution								
Hour	1	2	3	4	5	6	7	8
Employees needed	1	1	2	2	2	2	1	1
Employee A availability	Y	Y	Y	Y	Y	Y	Y	N
Employee B availability	N	N	Y	Y	Y	Y	Y	Y
Holistic schedule (one-phase)	Shift 1 (employee A)	W	W	W	W	W		
	Shift 2 (employee B)			W	W	W	W	W
	Total scheduled	1	1	2	2	2	2	1
	Net staffing level	0	0	0	0	0	0	0

Figura 2. 23: Confronto tra gli approcci ad una o due fasi per un problema di scheduling

2.4.4. Risoluzione dei problemi di scheduling

Possono essere usati diversi algoritmi per sviluppare una schedulazione della forza lavoro. Questi metodi possono essere distinti in ottimali e euristici. Una procedura ottimale consente di trovare la migliore soluzione possibile al problema di scheduling, ma, sfortunatamente, risulta efficace solo in contesti molto semplici. Per questa ragione le aziende operanti nel settore dell'ospitalità, caratterizzate da una realtà molto complessa, preferiscono utilizzare procedure euristiche, le quali sviluppano buone soluzioni in tempi abbastanza brevi.

Nello sviluppo delle schedulazioni devono inoltre essere fatte alcune considerazioni in merito a:

- Flessibilità
- Considerazioni sul personale
- Vincoli “hard” e “soft”
- Straordinario forzato e volontario.

Flessibilità

È di fondamentale importanza individuare le flessibilità da inserire nella schedulazione. Il vantaggio derivante da tale operazione consiste nella possibilità di adattare più facilmente il numero di impiegati schedulati al numero di addetti effettivamente necessari. Le flessibilità da individuare riguardano: la durata dei turni di lavoro, la collocazione delle pause, i tempi di inizio e fine di un turno, le mansioni affidate ad un singolo dipendente in uno specifico turno, l'esecuzione del lavoro controllabile. Quest'ultima forma assume particolare rilievo nel processo di scheduling.

La schedulazione del lavoro controllabile può essere svolta utilizzando un approccio a due fasi o un approccio a una fase. In figura 24[21] è mostrata un confronto tra tali metodi di schedulazione per un problema dove:

- l'orizzonte di scheduling è di 5 ore
- i periodi di pianificazione sono 5 da un'ora ciascuno
- i dipendenti sono addestrati a svolgere qualsiasi tipo di mansione
- i turni dei dipendenti hanno una durata di 3 ore
- le ore di lavoro controllabile da schedulare sono 3 in totale
- il numero ideale di impiegati necessari per svolgere il lavoro non controllabile è pari a tre unità per i periodi 2, 4, 5 ; quattro unità per il periodo 1 e cinque unità per il periodo 3.

Un approccio a due fasi tende a relegare il lavoro controllabile nei periodi in cui il lavoro non controllabile è minore, ossia nei periodi 2, 4 e 5. In questo modo si ottiene una schedulazione indicata nella parte superiore della figura. Tale schedulazione è costituita da 8 turni, di cui quattro iniziano nell'ora 1 e quattro nell'ora 3. Nell'ora 3 ci sono 8 dipendenti in servizio, il lavoro controllabile è già stato schedulato e il numero di staff necessario è pari a 5, quindi si registra la presenza di 3 impiegati inutilizzati.

Con un approccio a una fase invece si schedula il lavoro controllabile nei periodi in cui si ha un eccesso di dipendenti al lavoro. In tal modo si ottiene una schedulazione composta da 7 turni, di cui quattro cominciano nell'ora 1 e tre nell'ora 3. Il lavoro controllabile viene così schedulato: due ore nel periodo 3 e un'ora nel periodo 2.

Questa soluzione non prevede tempi morti dato che il personale schedulato è uguale a quello effettivamente necessario.

Appare evidente come la schedulazione del lavoro controllabile deve essere realizzata simultaneamente a quella del lavoro non controllabile per consentire il raggiungimento di obiettivi economici e di servizio.

Schedule with pre-assigned controllable work (D-framework solution)					
Hour	1	2	3	4	5
Employees needed for uncontrollable work	4	3	5	3	3
Assigned controllable work (total three hours)		1		1	1
Total employee requirements	4	4	5	4	4
Shift 1	w	w	w		
Shift 2	w	w	w		
Shift 3	w	w	w		
Shift 4	w	w	w		
Shift 5			w	w	w
Shift 6			w	w	w
Shift 7			w	w	w
Shift 8			w	w	w
Total scheduled employees	4	4	8	4	4
Overstaffing	0	0	3	0	0

Schedule resulting when controllable work is matched to employee availability (D-framework solution)					
Hour	1	2	3	4	5
Employees needed for uncontrollable work	4	3	5	3	3
Shift 1	w	w	w		
Shift 2	w	w	w		
Shift 3	w	w	w		
Shift 4	w	w	w		
Shift 5			w	w	w
Shift 6			w	w	w
Shift 7			w	w	w
Total scheduled employees	4	4	7	3	3
Scheduled controllable work		1	2		
Total workload	4	4	7	3	3
Overstaffing	0	0	0	0	0

Figura 2. 24: Confronto fra i metodi per la schedulazione del lavoro controllabile

Considerazioni sul personale

Nello sviluppo di una schedulazione vanno considerate le informazioni relative al personale. Esse si distinguono in considerazioni ambientali e considerazioni preferenziali.

Le considerazioni ambientali non sono determinate direttamente dai dipendenti. Esse includono: limitazioni sul numero minimo o massimo di ore lavorative giornaliere o settimanali, periodi di indisponibilità dei dipendenti, vincoli sul numero giorni di riposo consecutivi.

Le seconde rappresentano le preferenze dei dipendenti e possono comprendere: il numero totale di ore di lavoro giornaliere o settimanali, i giorni di riposo, l'assegnazione delle mansioni, la lunghezza della pausa pranzo, l'orario di inizio o di fine turno.

Entrambe le caratteristiche possono variare da un dipendente all'altro. Poichè raramente i dipendenti presentano caratteristiche preferenziali simili, un manager deve individuare differenze complementari che lo aiuteranno nella realizzazione dello scheduling.

Vincoli “hard” e “soft”

Solitamente i vincoli imposti rendono difficile lo sviluppo di un'ottima schedulazione. Essi possono essere classificati in due categorie: i vincoli “hard” e i vincoli “soft”.

I primi rappresentano quelle condizioni che non possono essere violate dalla schedulazione perché determinati da relazioni contrattuali o obbligatorie che l'azienda stabilisce con i dipendenti. Esempi tipici di vincoli rigidi sono quelli che definiscono il numero minimo o massimo accettabile di ore di lavoro giornaliere o settimanali, quelli che assicurano una corretta assegnazione dei dipendenti alle mansioni, quelli che schedulano i dipendenti solo quando sono questi sono effettivamente disponibili.

I vincoli “soft”, invece, sono quelli che devono essere soddisfatti, se possibile, ma non obbligatoriamente. Essi riguardano solitamente alcuni tipi di preferenze dei dipendenti.

Straordinario forzato o volontario

Le aziende del settore dell'ospitalità spesso operano in situazioni caratterizzate da un'elevata domanda, in cui i dipendenti in servizio non riescono a ricoprire il carico di lavoro richiesto durante l'orario di lavoro ordinario. Per questa ragione, i manager sono costretti a schedare turni di lavoro straordinario: forzato o volontario.

Da un punto di vista dello scheduling, lo straordinario forzato si realizza in modo molto semplice allungando la durata dei turni di lavoro o aggiungendo altri turni.

Schedare uno straordinario volontario risulta molto più complesso. Una possibile soluzione potrebbe essere quella di sviluppare dei turni non assegnati ad alcun impiegato e di renderli noti al personale per consentire ai dipendenti interessati di segnalarsi per ottenerne l'assegnazione.

Una delle difficoltà riscontrate nell'uso di tale metodo è legata al fatto che i dipendenti interessati ai turni potrebbero non avere le giuste competenze per il particolare turno.

E' possibile risolvere il problema raccogliendo le adesioni allo straordinario prima di effettuare la schedulazione in modo da assegnare gli straordinari a specifici dipendenti. La procedura da utilizzare è quella a una fase poichè solo in questo modo possono essere considerate tutte le limitazioni del caso evitando soluzioni non fattibili e realizzando obiettivi economici e di servizio.

2.5. Quarta fase: adattamento dello scheduling effettuato alla reale richiesta del servizio

Dopo la realizzazione delle prime tre fasi della procedura di scheduling, si avranno a disposizione: una previsione sui tassi di arrivo dei clienti, una lista del numero di dipendenti necessari e delle loro competenze, una descrizione dettagliata che specifica, per ogni dipendente in servizio, il luogo e la durata del turno di lavoro.

Le prime tre fasi sono tutte attività di previsione svolte antecedentemente all'effettiva distribuzione del servizio. Al contrario, l'ultima fase riguarda la regolazione della schedulazione in tempo reale. In essa, si valuta se la soluzione scaturita dai pronostici garantisce realmente che i clienti siano adeguatamente serviti. Questa fase, che consiste nel confronto tra la realtà operativa e la schedulazione pianificata, è l'essenziale pezzo finale che garantisce un soddisfacente servizio per i clienti. Spesso, infatti, esiste uno squilibrio tra la capacità di lavoro e la richiesta del servizio dovuta a diversi fattori, tra i quali la mancata corrispondenza tra richiesta del servizio prevista e quella effettiva e l'incertezza sulle prestazioni realmente fornite dal personale (alcuni dipendenti potrebbero essere malati o in ritardo).

E' di fondamentale importanza non solo valutare se la schedulazione prevista si adatta di fatto alla richiesta reale dei consumatori, ma anche adottare particolari provvedimenti per realizzare tale corrispondenza.

Azioni di regolazione in tempo reale

Le azioni di regolazione in tempo reale possono essere distinte in:

- Azioni di breve durata
- Azioni di lunga durata.

Le prime hanno effetto solo su un breve periodo di un giorno operativo (tipicamente da qualche minuto fino ad un'ora) e sono facilmente revocabili. Esempi di tali azioni sono: mandare i dipendenti in pausa o richiamarli dalla pausa, estendere la lunghezza di un turno, chiedere ai dipendenti di svolgere mansioni differenti per brevi momenti.

Le seconde hanno effetti su periodi più lunghi di un'ora e richiedono un maggiore impiego di risorse. Alcuni esempi sono: mandare i dipendenti a casa in anticipo, chiamare in servizio dipendenti aggiuntivi, riassegnare i dipendenti a mansioni diverse.

Il problema chiave della regolazione in tempo reale è stabilire quando compiere un'azione che modifica la schedulazione originale e se intraprendere un'azione di breve o lunga durata.

Le azioni di breve durata hanno effetti relativamente piccoli sui costi e sul servizio, mentre quelle di lunga durata non solo possono influenzare le caratteristiche del servizio, ma, una volta intraprese, sono difficilmente reversibili. Solo quando un manager è in grado di prevedere la reale richiesta di servizio, stabilendo se è maggiore o minore di quella prevista, può intraprendere con fiducia azioni di lunga durata, in caso contrario egli dovrà limitarsi a stabilire solo azioni di breve durata.

2.5.1. Un approccio per prevedere la richiesta del servizio in un dato giorno

Di seguito viene proposto un metodo che consente di prevedere l'affluenza di clienti in un dato giorno. Esso si compone dei seguenti cinque passi:

- 1) Determinare la consistenza nei dati rappresentativi della richiesta del servizio
- 2) Identificare la porzione di vendite derivante da ogni periodo di pianificazione in un dato orizzonte temporale
- 3) Classificare ogni giorno in base al suo volume di business
- 4) Eseguire delle simulazioni per sviluppare diagrammi delle realizzazioni per ogni livello del volume di business
- 5) Realizzare una simulazione per prevedere il reale volume di business giornaliero.

Step 1

Il primo passo di tale approccio consiste nella determinazione della consistenza dei dati. Il procedimento, analogo a quello illustrato nella prima fase dello scheduling, consiste nella iniziale raccolta di un certo numero di dati in un determinato orizzonte di riferimento e nella successiva realizzazione di un'analisi per determinare se tra essi sussiste una perfetta correlazione.

In figura 25[22] vengono rappresentate le vendite realizzate in ogni periodo di pianificazione come percentuale delle vendite giornaliere totali, considerando un orizzonte di riferimento di un giorno e periodi di pianificazione di 15 minuti. La raccolta dei dati prende in considerazione come giorno il lunedì, e si estende per 4 settimane consecutive.

Il diagramma mette in evidenza che:

- Le curve rappresentative delle vendite hanno un andamento simile nei quattro lunedì consecutivi
- Si registra la presenza di un picco secondario intorno al periodo 15, una fase di arresto intorno al periodo 25 e un picco primario intorno al periodo 40 seguito da un abbassamento nel resto del giorno.

Confrontando i dati raccolti attraverso un'analisi di correlazione, riportata in figura 26[22], si evince la consistenza richiesta nei dati, grazie alla presenza di un valore dell'indice ovunque maggiore di 0.79.

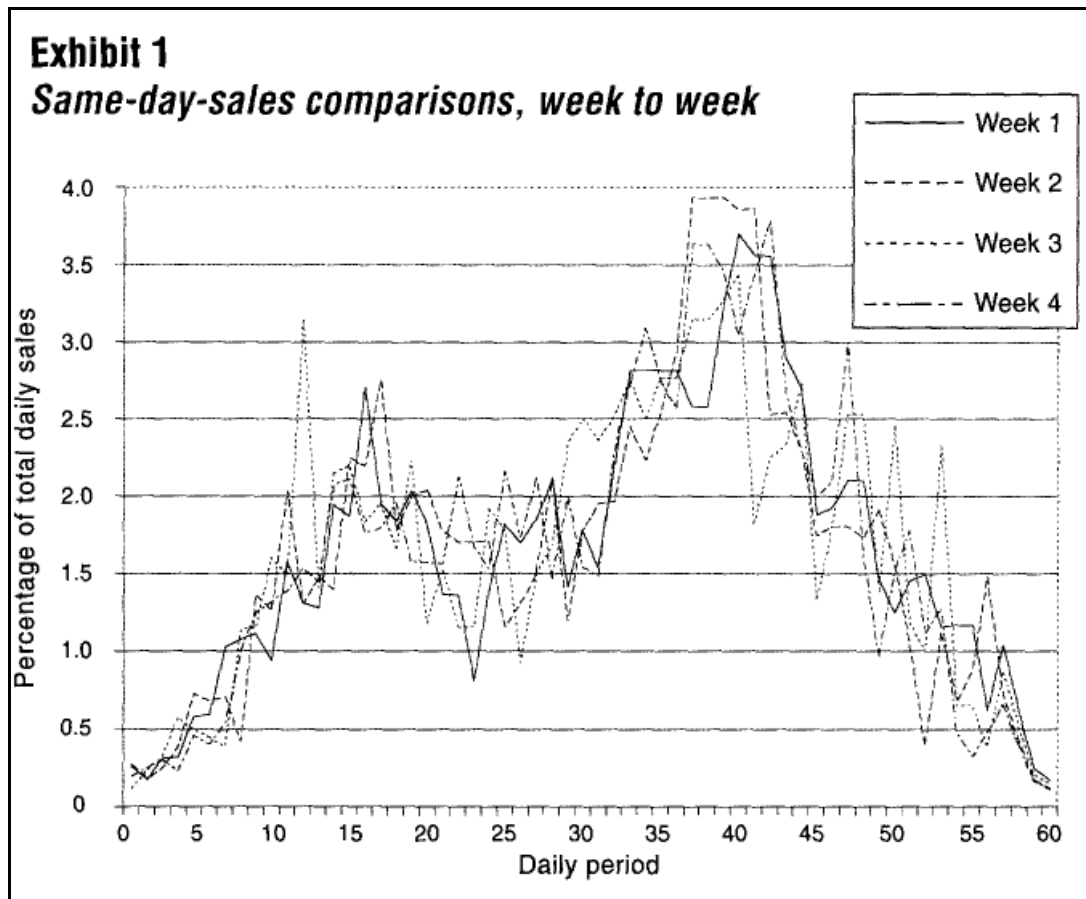


Figura 2.25: Confronto tra le vendite in quattro lunedì consecutivi

Exhibit 2
Correlations between weeks for the data in Exhibit 6

	Week 1	Week 2	Week 3	Week 4
Week 1	1			
Week 2	0.8708	1		
Week 3	0.7954	0.7889	1	
Week 4	0.9039	0.8679	0.8262	1

Figura 2.26: Analisi di correlazione

Step 2

Una volta stabilita la consistenza nei dati, il manager deve identificare le proporzioni delle vendite giornaliere in ogni periodo di pianificazione (solitamente di durata pari a 15 minuti), provvedendo inoltre a ridurre la variazione casuale mediante l'applicazione di tecniche di smoothing che producono un appianamento dei picchi e delle vallate presenti nei diagrammi. In figura 27[22] è rappresentato un tipico risultato dell'applicazione di questa tecnica.

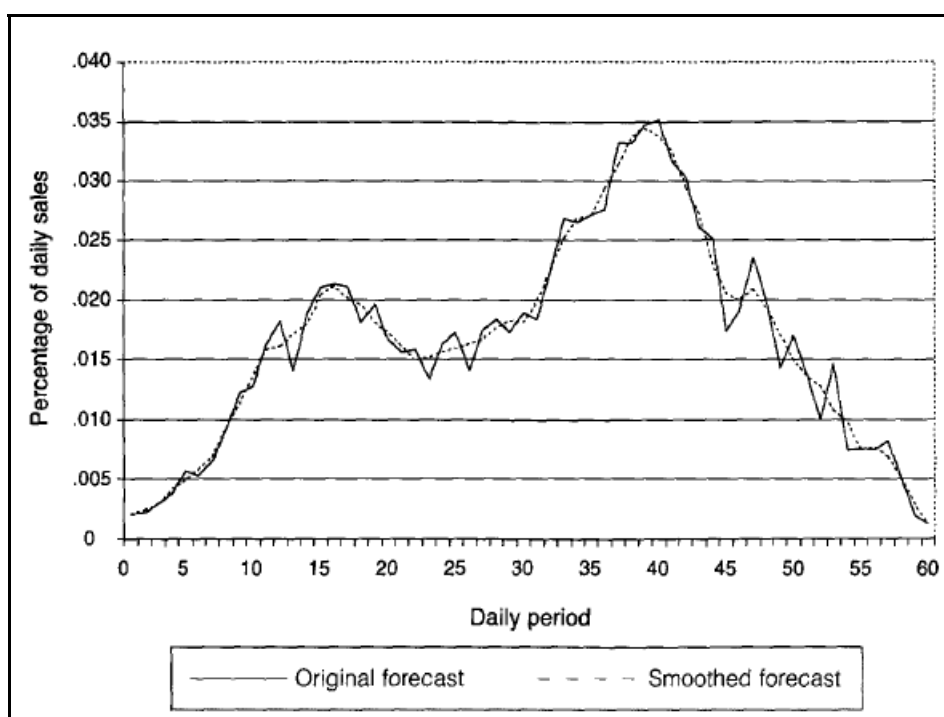


Figura 2. 27: Risultato dell'applicazione della tecnica di smoothing

Step 3

Un metodo razionale per classificare un giorno in base al suo volume di attività è costituito dalla suddivisione dell'intervallo rappresentativo del volume totale di clienti da servire nel giorno considerato in cinque intervalli, corrispondenti a cinque diversi volumi di business. Questi intervalli sono detti "livelli". Il livello 1 indica il più basso volume di business, il livello 5 il più alto. Si prenda in considerazione, ad esempio, un'azienda del settore dell'ospitalità che serve tipicamente, in un dato giorno, un flusso di

clienti variabile da 500 a 1500. Suddividendo tale intervallo in cinque categorie si determineranno i seguenti livelli:

- Livello 1: corrispondente a un flusso di clienti che varia da 500 a 700
- Livello 2: corrispondente a un flusso di clienti che varia da 701 a 900
- Livello 3: corrispondente a un flusso di clienti che varia da 901 a 1100
- Livello 4: corrispondente a un flusso di clienti che varia da 1101 a 1300
- Livello 5: corrispondente a un flusso di clienti che varia da 1301 a 1500.

Per ognuna di tali categorie un manager può calcolare gli arrivi dei clienti in ogni periodo di pianificazione in cui è suddiviso l'orizzonte di riferimento, ossia il giorno, ottenendo un diagramma chiamato "realizzazione" e mostrato in figura 28[22].

Dall'analisi di tale "realizzazione" è possibile notare come i dati rappresentati si mostrino coerenti con quelli raccolti in figura 26, registrando un picco secondario intorno al periodo 15, una vallata in corrispondenza del periodo 25 e, infine, un picco primario in prossimità del periodo 40.

Uno strumento più utile della realizzazione del volume di business nei vari periodi è la realizzazione del volume di business cumulativa mostrata in figura 29[22] relativa agli stessi dati utilizzati per la precedente realizzazione. In essa viene messo in evidenza il numero complessivo di clienti serviti fino a un determinato periodo. Ad esempio, considerando il diagramma, si può affermare che sono stati serviti circa 100 clienti fino al periodo 15, circa 200 fino al periodo 26 e circa 300 fino al periodo 34.

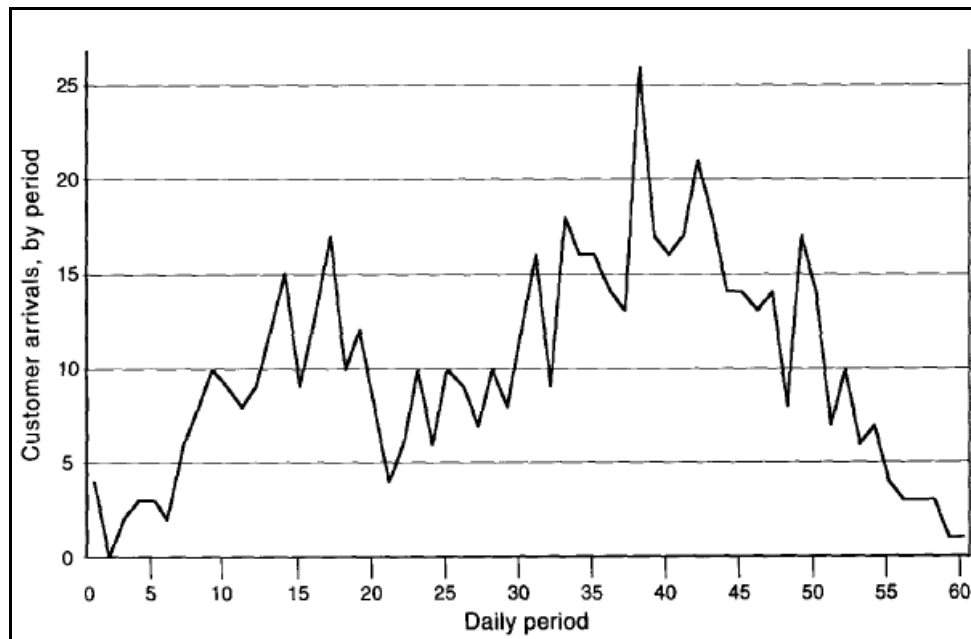


Figura 2.28: Realizzazione nei periodi corrispondente al livello 1

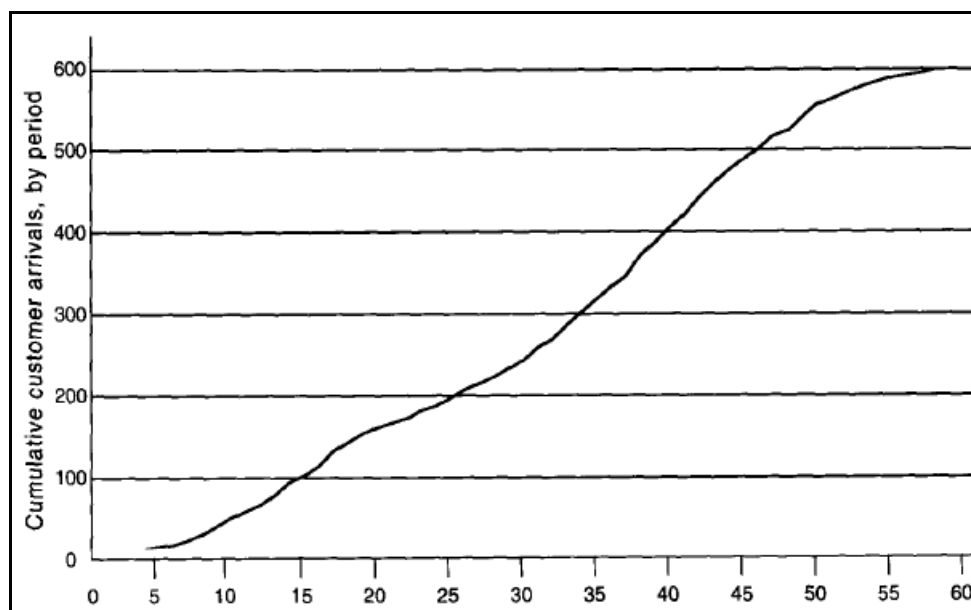


Figura 2.29: Realizzazione del volume di business cumulativa

Step 4

Per realizzare l'aggiustamento tra la domanda reale e quella effettiva è necessario avere a disposizione un set di "realizzazioni" cumulative con le quali confrontare un dato campione di domanda giornaliera. È necessario a tal fine produrre un set di circa 200 realizzazioni per ogni livello di volume di business determinato, utilizzando la simulazione come strumento per la sua generazione. Queste realizzazioni si ottengono applicando un'equazione ricavata considerando che gli arrivi dei clienti seguono una distribuzione di Poisson, ossia:

$$t + [(-1 \div n) \times \ln(R)]$$

Dove:

- n rappresenta il tasso medio di arrivo dei clienti
- $\ln(R)$ rappresenta il logaritmo naturale di un numero casuale, R , compreso nell'intervallo $[0,1]$
- t rappresenta il periodo in cui si registra l'arrivo di un cliente.

In figura 30[22], delle 200 realizzazioni effettuate per il livello 1, considerando la variazione del tasso di arrivo di clienti nel corso della giornata (da 500 a 700 arrivi), ne sono plottate soltanto cinque ossia:

- linea al 100%: identifica il maggior numero di clienti servito in qualunque periodo del giorno
- linea allo 0%: identifica il minor numero di clienti servito in qualunque periodo del giorno
- linea al 50%: identifica il numero medio di clienti servito in qualunque periodo del giorno
- linea al 25%: identifica il primo quartile del conteggio di clienti
- linea al 75%: identifica il terzo quartile del conteggio di clienti.

Osservando il periodo 30 in tale figura si può notare come non sono stati serviti mai più di 307 clienti (linea 100%) e mai meno di 188 (linea 0%), quindi in media 250 clienti (linea 50%).

In figura 31[22] sono plottate le stesse realizzazioni per il livello 2, corrispondente ad un intervallo di arrivi compreso tra 701 e 900. Considerando ancora il periodo 30, si osserva che non sono mai stati serviti più di 405 clienti (linea 100%) e mai meno di 272 (linea 0%), con una media di 334 clienti circa (linea 50%).

Le realizzazioni corrispondenti ai diversi volumi di business effettuate attraverso la simulazione talvolta si accavallano. Ad esempio, supponendo che nel periodo 30 sono stati serviti 290 clienti si può notare come questo valore appartenga sia all'intervallo di clienti serviti nel periodo considerato corrispondente al livello 1 che a quello relativo al livello 2.

Risulta importante, in tal caso, per un manager saper individuare opportunamente il livello cui appartiene il conteggio di clienti in un dato giorno.

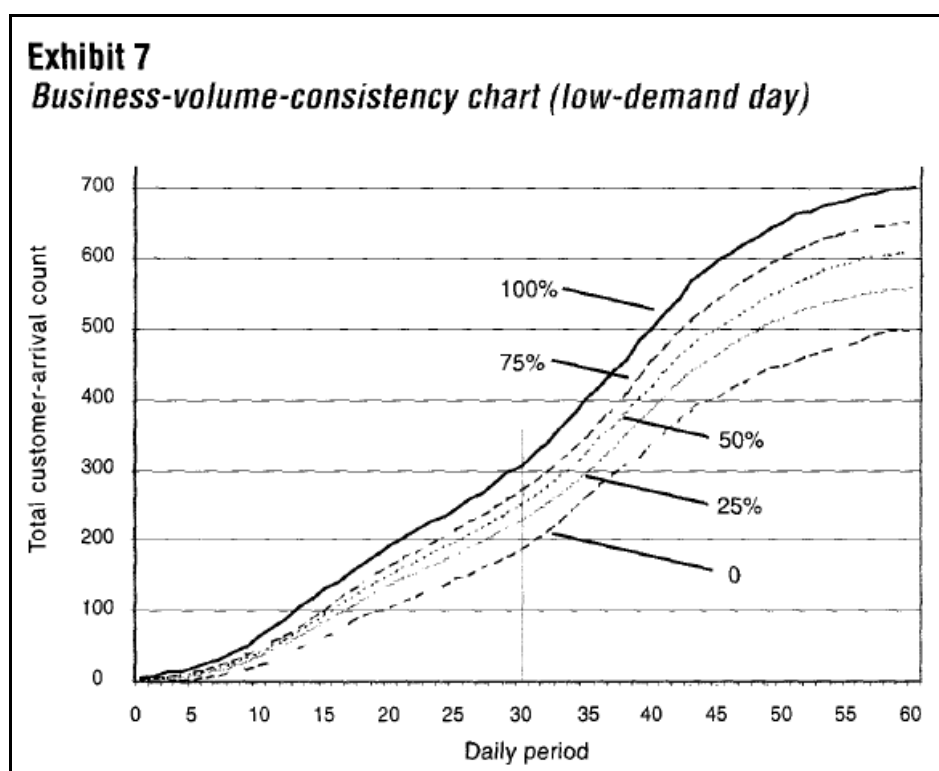


Figura 2.30: Diagramma delle realizzazioni per il livello 1 del volume di Business

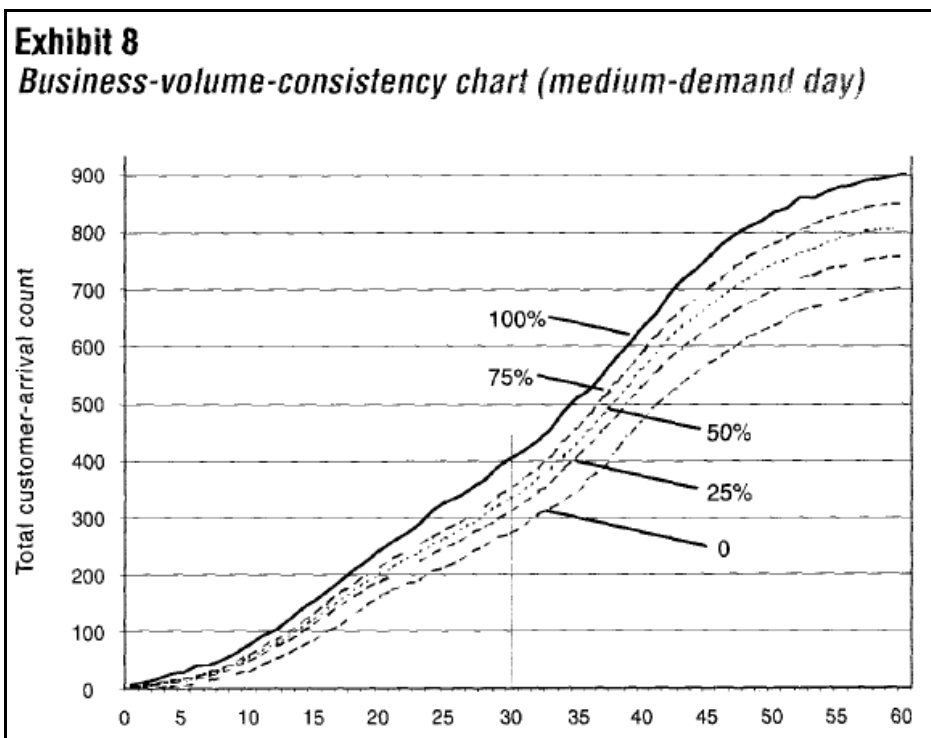


Figura 2.31: Diagramma delle realizzazioni per il livello 2 del volume di business

Step 5

Per determinare con precisione a quale livello di business appartiene la realizzazione ottenuta in un dato giorno, il manager può eseguire un conteggio cumulativo dei clienti all'inizio del giorno ed effettuare una simulazione per ottenere la "realizzazione" corrispondente al conteggio cumulativo dei clienti durante tutto il giorno.

Successivamente egli deve conteggiare il numero totale di realizzazioni confrontabili e valutare, in percentuale, l'incidenza su tale numero esercitata dalle realizzazioni relative ad ogni livello di business stabilito. Alla percentuale con valore maggiore corrisponderà il livello di business dove includere la realizzazione del giorno corrente.

Per realizzazione confrontabile si intende una realizzazione in cui il conteggio cumulativo di clienti nel periodo (i-1) è uguale o minore rispetto al valore del conteggio nel periodo (i) della realizzazione reale e il conteggio nel periodo (i+1) è maggiore o uguale rispetto al conteggio (i) della realizzazione reale.

Un esempio di questa procedura è mostrato in figura 32[22]: alla fine del periodo 10 è stato contato l'arrivo di 45 clienti, delle 200 realizzazioni effettuate per il livello 1, solo 99 sono risultate confrontabili con la realizzazione determinata partendo dal tasso di arrivo registrato in tale periodo. Il livello 2 ha totalizzato 30 realizzazioni confrontabili, il livello 3 ne ha registrate 3, infine, i livelli 4 e 5 nessuna.

Il numero totale di realizzazioni confrontabili è dato da:

$$99 + 30 + 3 = 132$$

La percentuale di queste determinata dai singoli livelli di business è la seguente:

- livello 1: 75%
- livello 2: 22.73%
- livello 3: 2.27%.

Nel periodo 10 quindi c'è una probabilità del 75% che il conteggio di clienti ricada nel livello 1.

Considerando poi il periodo 13 è possibile verificare che tale probabilità ancora più forte, dato il valore della percentuale pari a 92.50.

Dai dati rappresentati in figura 32 può essere generato il diagramma in figura 33[22].

Esso mostra come in corrispondenza del periodo 1 si può rilevare una probabilità del 25.56% che il conteggio ricada nel livello 1, del 23.99% che ricada nel livello 2, e così via.

Naturalmente, più grande è il valore della probabilità, più facilmente si potrà realizzare un'accurata previsione del livello di business in cui inserire il conteggio reale di clienti.

I primi quattro step della procedura possono essere svolti periodicamente, l'ultimo step invece dovrebbe essere eseguito ogni ora o anche più spesso, poichè esso consente il monitoraggio del volume di affari giornaliero.

Period	Cumulative Actual Customer Count	Number of Comparable Cases (percentage of total comparable cases)					Total Comparable Realizations
		Level 1	Level 2	Level 3	Level 4	Level 5	
1	2	68 (25.56%)	69 (25.94%)	51 (19.17%)	43 (16.17%)	35 (13.16%)	266
2	7	185 (23.99%)	179 (23.22%)	150 (19.46%)	145 (18.81%)	112 (14.53%)	771
3	10	71 (16.55%)	100 (23.31%)	94 (21.91%)	94 (21.91%)	70 (16.32%)	429
4	13	63 (18.16%)	96 (27.67%)	78 (22.48%)	74 (21.33%)	36 (10.37%)	347
5	14	39 (31.97%)	46 (37.7%)	21 (17.21%)	12 (9.84%)	4 (3.28%)	122
6	21	117 (35.67%)	119 (36.28%)	62 (18.9%)	26 (7.93%)	4 (1.22%)	328
7	25	63 (35.80%)	74 (42.05%)	33 (18.75%)	6 (3.41%)	0 (0)	176
8	28	52 (61.90%)	28 (33.33%)	4 (4.76%)	0 (0)	0 (0)	84
9	37	115 (69.28%)	48 (28.92%)	3 (1.81%)	0 (0)	0 (0)	166
10	45	99 (75.00%)	30 (22.73%)	3 (2.27%)	0 (0)	0 (0)	132
11	53	93 (83.78%)	18 (16.22%)	0 (0)	0 (0)	0 (0)	111
12	65	119 (87.50%)	17 (12.5%)	0 (0)	0 (0)	0 (0)	136
13	73	74 (92.50%)	6 (7.5%)	0 (0)	0 (0)	0 (0)	80
14	87	102 (92.73%)	8 (7.27%)	0 (0)	0 (0)	0 (0)	110
15	103	101 (89.38%)	12 (10.62%)	0 (0)	0 (0)	0 (0)	113
16	116	73 (91.25%)	7 (8.75%)	0 (0)	0 (0)	0 (0)	80
17	128	62 (95.38%)	3 (4.62%)	0 (0)	0 (0)	0 (0)	65
18	147	82 (90.11%)	9 (9.89%)	0 (0)	0 (0)	0 (0)	91
19	155	33 (89.19%)	4 (10.81%)	0 (0)	0 (0)	0 (0)	37
20	171	54 (85.71%)	9 (14.29%)	0 (0)	0 (0)	0 (0)	63
21	186	38 (74.51%)	13 (25.49%)	0 (0)	0 (0)	0 (0)	51
22	198	32 (68.09%)	15 (31.91%)	0 (0)	0 (0)	0 (0)	47
23	204	23 (85.19%)	4 (14.81%)	0 (0)	0 (0)	0 (0)	27
24	215	33 (86.84%)	5 (13.16%)	0 (0)	0 (0)	0 (0)	38
25	221	14 (100)	0 (0)	0 (0)	0 (0)	0 (0)	14
26	229	22 (95.65%)	1 (4.35%)	0 (0)	0 (0)	0 (0)	23
27	239	33 (100)	0 (0)	0 (0)	0 (0)	0 (0)	33
28	248	27 (100)	0 (0)	0 (0)	0 (0)	0 (0)	27
29	259	30 (100)	0 (0)	0 (0)	0 (0)	0 (0)	30
30	270	27 (100)	0 (0)	0 (0)	0 (0)	0 (0)	27
31	288	42 (97.67%)	1 (2.33%)	0 (0)	0 (0)	0 (0)	43
32	307	41 (93.18%)	3 (6.82%)	0 (0)	0 (0)	0 (0)	44
33	332	48 (92.31%)	4 (7.69%)	0 (0)	0 (0)	0 (0)	52
34	341	21 (95.45%)	1 (4.55%)	0 (0)	0 (0)	0 (0)	22
35	352	23 (92%)	2 (8%)	0 (0)	0 (0)	0 (0)	25
36	368	35 (100)	0 (0)	0 (0)	0 (0)	0 (0)	35
37	387	40 (100)	0 (0)	0 (0)	0 (0)	0 (0)	40
38	412	50 (100)	0 (0)	0 (0)	0 (0)	0 (0)	50
39	443	53 (100)	0 (0)	0 (0)	0 (0)	0 (0)	53
40	464	38 (100)	0 (0)	0 (0)	0 (0)	0 (0)	38
41	485	31 (100)	0 (0)	0 (0)	0 (0)	0 (0)	31
42	498	21 (100)	0 (0)	0 (0)	0 (0)	0 (0)	21
43	513	24 (100)	0 (0)	0 (0)	0 (0)	0 (0)	24
44	531	27 (100)	0 (0)	0 (0)	0 (0)	0 (0)	27
45	543	15 (100)	0 (0)	0 (0)	0 (0)	0 (0)	15
46	550	8 (100)	0 (0)	0 (0)	0 (0)	0 (0)	8
47	557	10 (100)	0 (0)	0 (0)	0 (0)	0 (0)	10
48	571	20 (100)	0 (0)	0 (0)	0 (0)	0 (0)	20
49	576	7 (100)	0 (0)	0 (0)	0 (0)	0 (0)	7
50	582	6 (100)	0 (0)	0 (0)	0 (0)	0 (0)	6
51	594	12 (100)	0 (0)	0 (0)	0 (0)	0 (0)	12
52	604	11 (100)	0 (0)	0 (0)	0 (0)	0 (0)	11
53	607	3 (100)	0 (0)	0 (0)	0 (0)	0 (0)	3
54	615	7 (100)	0 (0)	0 (0)	0 (0)	0 (0)	7
55	620	5 (100)	0 (0)	0 (0)	0 (0)	0 (0)	5
56	624	3 (100)	0 (0)	0 (0)	0 (0)	0 (0)	3
57	628	3 (100)	0 (0)	0 (0)	0 (0)	0 (0)	3
58	633	3 (100)	0 (0)	0 (0)	0 (0)	0 (0)	3
59	636	5 (100)	0 (0)	0 (0)	0 (0)	0 (0)	5
60	636	3 (100)	0 (0)	0 (0)	0 (0)	0 (0)	3

Figura 2. 32: Procedura per la determinazione del livello di business cui corrisponde la domanda reale

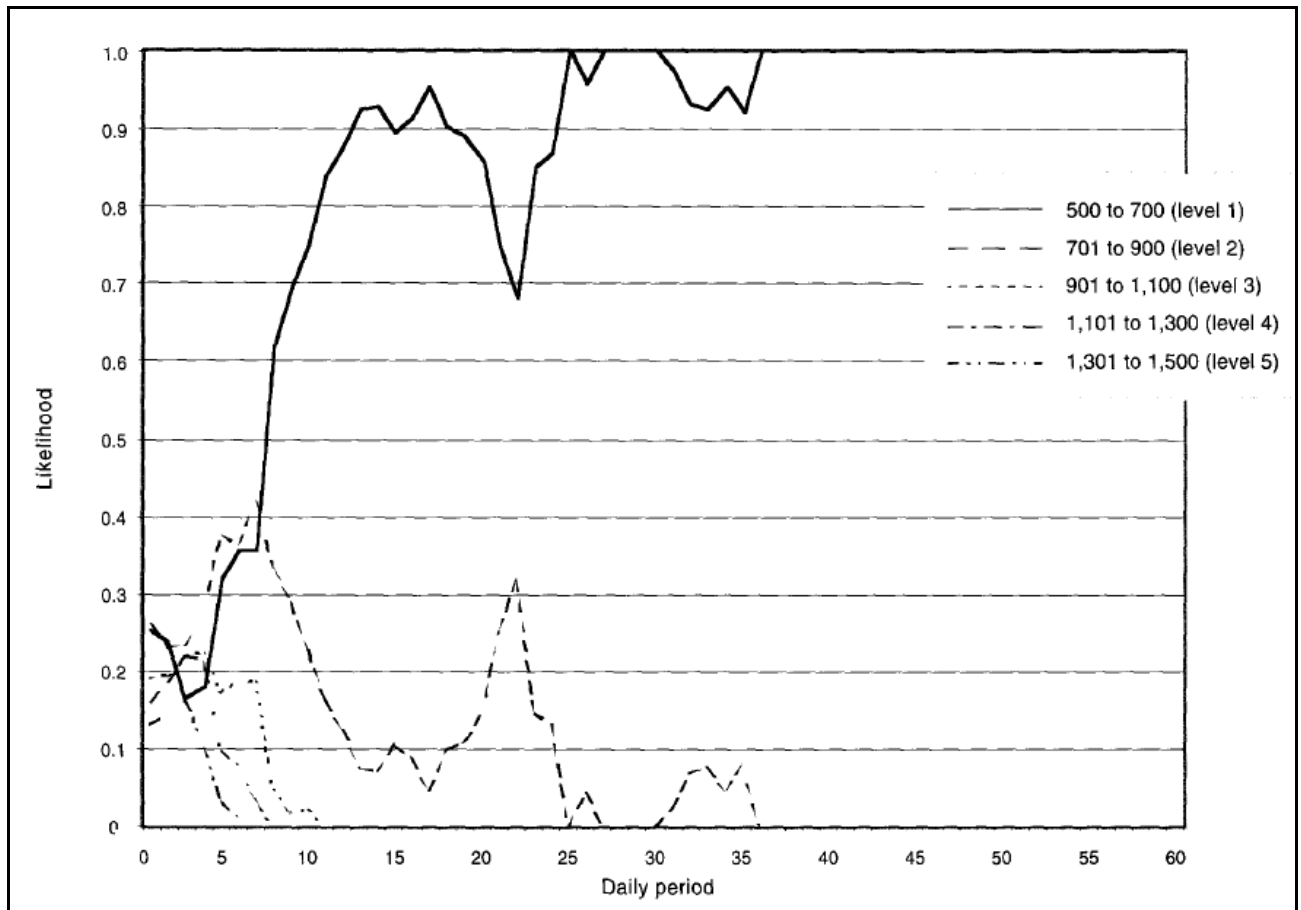


Figura 2.33: Probabilità che il conteggio di clienti corrispondente ad un determinato periodo ricada nei diversi livelli di business

2.5.2. Azioni da sviluppare in tempo reale e importanza del “cross-training”

Prevedendo in anticipo il volume giornaliero di business il manager può intraprendere azioni di lunga durata per realizzare le opportune modifiche alla schedulazione del personale ottenuta dalle prime tre fasi dello scheduling.

Ad esempio, nel caso in cui il manager possiede un indicazione non molto forte su un volume di business pari a un livello 3, egli tenderà a sviluppare una schedulazione basata sul livello 4 per ottenere la garanzia che ogni cliente riceverà un servizio appropriato.

Nel caso in cui siano presenti forti indicazioni sul volume di business, sarà possibile intraprendere azioni di lunga durata come:

- chiedere a qualche addetto di rinunciare a un giorno di lavoro, nel caso il volume di business corrisponda a quello più basso
- chiamare dipendenti extra o imporre lo straordinario, quando avendo stabilito una schedulazione sulla base del livello 4 si avranno forti segnali che il reale volume di attività corrisponde al livello 5.

Quando non ci sono indicazioni chiare sul volume di business giornaliero sarà più opportuno realizzare solo azioni di controllo di breve durata.

Anche una previsione relativamente solida può contenere possibilità di errore. Un modo per ridurre tale incertezza è assumere dipendenti “cross-trained”, guadagnando flessibilità nella schedulazione grazie all’opportunità di poter collocare i dipendenti dove ce n’è più bisogno.

Il valore legato ai dipendenti cross-trained può essere meglio compreso considerando il seguente esempio. Si supponga che un attività sia composta da tre mansioni e che non ci sia la possibilità di utilizzare dipendenti cross-trained. Attraverso una previsione accurata (caratterizzata da un coefficiente di variazione dell’errore basso, ad esempio $COV = 0.05$) si determina che ogni mansione debba essere ricoperta da 10 dipendenti generando un costo orario totale del lavoro pari a \$335.71.

Con l’aumento dell’incertezza della previsione ($COV = 0.25$) il livello di staff determinato per ogni attività è pari a 13 unità corrispondente ad un valore di costo di \$403.81. Questa situazione è mostrata nella parte a sinistra della figura 34[22]. Nella parte destra viene mostrato il vantaggio derivante dall’uso di dipendenti “cross-trained”. Nel primo scenario illustrato i dipendenti cross-trained ricevono una salario orario maggiore di circa il 20% rispetto al salario ordinario. La situazione ideale corrispondente al valore $COV = 0.25$, nel primo scenario, è data dall’assegnazione di 9 dipendenti standard per ogni mansione e di 6 dipendenti cross-trained. Tale disposizione del lavoro richiede un numero totale di dipendenti pari a 33 a un costo di \$371.54, ossia comporta una riduzione del numero di addetti pari al 15% e dei costi pari all’8% rispetto alla situazione in cui manca la possibilità di adoperare dipendenti cross-trained. I risparmi conseguiti sono naturalmente ancora maggiori nel secondo scenario, in figura 32[22], dove la paga oraria di questi dipendenti è maggiore rispetto quella dei dipendenti ordinari solo del 10%.

In definitiva quando è presente una maggiore incertezza nella domanda usare dipendenti cross-trained può portare vantaggi legati all’opportunità di contenere sia i costi che il numero di dipendenti.

	base case		scenario 1 (20% wage & benefit premium			
	(no cross-trained employees)		for cross-trained employees)			
COV	staffing level	hourly cost	staffing level	hourly cost	hourly savings	% savings
0,25	13/13/13/0	\$ 403,81	9/9/9/6	\$ 371,54	\$ 32,27	7,99
0,2	12/12/12/0	\$ 373,51	10/10/10/3	\$ 360,41	\$ 18,10	4,78
0,15	11/11/11/0	\$ 352,69	10/10/10/2	\$ 349,77	\$ 7,82	2,19
0,1	10/10/10/0	\$ 347,62	10/10/10/1	\$ 341,26	\$ 6,36	1,83
0,05	10/10/10/0	\$ 335,71	10/10/10/0	\$ 335,71	\$ 0	0
0	10/10/10/0	\$ 333,12	10/10/10/0	\$ 333,12	\$ 0	0

Figura 2.34: Flessibilità fornita dall'utilizzo dei dipendenti cross-trained

	scenario 2 (10% wage & benefit premium			
	for cross-trained employees)			
COV	staffing level	hourly cost	hourly savings	% savings
0,25	9/9/9/6	\$ 365,54	\$ 38,27	9,48
0,2	9/9/9/5	\$ 355,41	\$ 23,10	6,1
0,15	9/9/9/4	\$ 346,69	\$ 10,90	3,05
0,1	10/10/10/1	\$ 340,26	\$ 7,36	2,12
0,05	10/10/10/0	\$ 335,71	\$ 0	0
0	10/10/10/0	\$ 333,12	\$ 0	0

Figura 2.35: Flessibilità fornita dall'utilizzo dei dipendenti cross-trained

CAPITOLO III

IL BILANCIO DELLA CAPACITA' PRODUTTIVA IN FRONT OFFICE: LA GESTIONE DELLE CODE

Nei precedenti capitoli, dopo aver introdotto il concetto di sistema di servizio, si è posta particolare attenzione a quelle che sono le problema che connesse con una corretta gestione della loro capacità produttiva.

In particolare si è evidenziato che, essendo la caratteristica principale di un sistema di servizio l'impossibilità di stoccaggio, diventa fondamentale da una lato la gestione delle risorse umane e dall'altro lo studio della teoria delle code relativa ai clienti.

In particolare la gestione delle risorse umane avviene attraverso specifici algoritmi per lo scheduling del personale ed è funzionale alle richieste del cliente e quindi alla corretta gestione delle code di attesa dei clienti.

Spesso nella vita quotidiana, si ha a che fare con delle situazioni in cui viene richiesto un servizio e dove vi sono delle unità adibite a tale scopo. I sistemi di servizio in cui ciò avviene sono, di solito, soggetti ad aleatorietà e ciò comporta problemi di congestione, la cui gestione è oggetto di analisi della Teoria delle Code. Essa è una disciplina che ha oltre cento anni e che ingloba in sé concetti, tecniche, metodi di diverse altre discipline, applicate con l'obiettivo di allievare la coda (e le conseguenze che essa comporta) ed è utilizzata in diversi ambiti, oltre a quello di servizio su cui è incentrato tale elaborato di tesi.

*In questo capitolo si descrivono le componenti di cui è costituito un sistema di servizio e le quantità da tenere in considerazione per una **analisi stazionaria** di un sistema di code e saranno presi in considerazione i **processi di arrivo e di servizio** che influenzano l'andamento dei clienti in arrivo e in uscita dal sistema ed, in particolare, quelli di tipo Markoviano, che meglio rappresentano la realtà.*

*Infine, verrà menzionato un metodo utile per studiare le code, ovvero, il **processo di nascita e morte**, evidenziandone l'applicazione delle sue proprietà generali per la determinazione delle misure di prestazione nel caso dei sistemi di code più significativi.*

3.1 Centri Di Contatto E Fenomeni Di Congestione

Grazie allo sviluppo tecnologico che ha interessato le aziende dei servizi, l'interazione tra clienti, che richiedono un servizio, e serventi, che soddisfano tale esigenza, può avvenire tramite web, posta elettronica, fax, telefono, messengerie su telefoni cellulari, registrazioni vocali (VoIP) e non solo attraverso uno "sportello fisico".

I **contact centers** o **centri di contatto** sono un insieme di risorse che forniscono un'interfaccia tra il fornitore di servizi e i propri clienti [2], attraverso qualsiasi canale o mezzo.

Un centro di contatto può essere considerato lo scenario in cui si concretizzano i, cosiddetti, **sistemi di servizio**, ossia l'insieme di elementi, quali clienti e server, che interagiscono tra loro per il raggiungimento di uno scopo comune, quale la realizzazione del servizio. L'arrivo dei clienti, solitamente, è di tipo casuale ed ognuno richiede un servizio la cui realizzazione avviene con una certa durata, variabile caso per caso. Nasce, con queste basi, una *coda* costituita da clienti che attendono di essere serviti.

In ambito dei servizi, essendo i protagonisti della coda uomini e non beni materiali (come, ad esempio, può avvenire in ambito produttivo), il **problema della congestione** comporta disagi e disagi vari.

La **teoria delle code in ambito dei servizi**, si occupa di analizzare tali fenomeni di attesa modellando il comportamento e il processo di arrivo di un cliente (domanda di servizio), e le modalità e il processo con cui viene erogato tale servizio dai serventi (offerta di servizio).

Come in ogni situazione in cui è presente un'offerta a fronte di una domanda, ci si pone come obiettivo di trovare e di raggiungere un punto di equilibrio tra esigenze contrastanti:

- i clienti desiderano ricevere il servizio prima possibile
- il gestore del servizio deve dimensionare il sistema in modo da massimizzare il livello di soddisfazione dei propri clienti (*costi della coda*) e, contemporaneamente, minimizzare i costi sostenuti nel fornire il servizio, principalmente, *costi del personale*.

La teoria delle code, attraverso un modello (descrittivo o simulativo), studia il sistema di code e ricerca il numero di serventi ottimale in corrispondenza del costo minimo totale (somma di costi dovuti all'attesa in coda e costi del personale), equilibrando gli obiettivi sopra citati (**Fig. 1**).

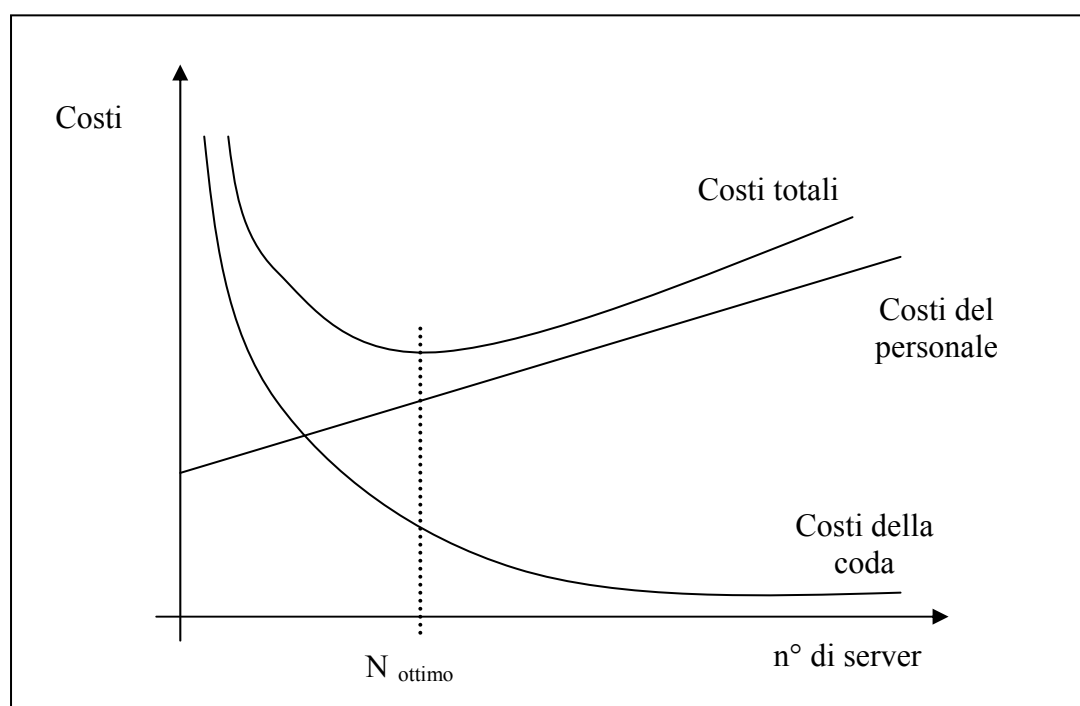


Figura 3.1. Dimensionamento ottimale di una stazione di servizio.

Nella maggior parte dei casi, come già detto, la domanda di servizio e i tempi di espletamento del servizio sono variabili, per cui è forte l'esigenza di tale dimensionamento del sistema e di un'ottima coordinazione

tra le parti che lo costituiscono. Ed in questo, la teoria delle code risulta essere un ottimo strumento di pianificazione e gestione di un sistema di servizio.

3.2 I Passi Fondamentali Nella Teoria Delle Code

La teoria delle code ha origine oltre cento anni fa. Il primo approccio allo studio di una coda in un sistema risale al 1907 con l'articolo "*Waiting times and number of calls*" di Johannsen. Il metodo utilizzato, però, risulta non essere matematicamente esatto, per cui come vero fautore della *Teoria delle code* viene considerato l'ingegnere danese A.K.Erlang che nel 1909 presentò, con "*The theory of probabilities and telephone conversations*", il primo studio dettagliato sul traffico telefonico, introducendo la distribuzione di Poisson e, quindi esponenziale (vedi par. 1.5), nello studio di un sistema di code. Erlang nei successivi venti anni diede un grande contributo alla teoria delle code. Nei suoi lavori propose concetti e tecniche di una certa rilevanza, di cui il metodo delle equazioni di stato (in seguito denominato equazioni di Chapman-Kolmogorov) e la nozione di equilibrio statistico ne rappresentano due esempi [3].

Nelle successive due decadi, diversi teorici si occuparono di fenomeni di congestione, sviluppando modelli generali che potessero essere usati in diverse e complesse situazioni. Tra i più importanti citiamo Feller e Kolmogorov che, studiando processi discontinui, gettarono le basi alla teoria dei processi di Markov. Pollaczek [4], invece, analizzò sistemi di code non all'equilibrio e osservò il loro comportamento in un intervallo di tempo finito. Le prime soluzioni di problemi a tempo dipendente (analisi transiente) furono proposte da Bailey [5], che utilizzò funzioni generatrici per le equazioni differenziali, e Lederman e Reuter [6] che, invece, usarono la Teoria Spettrale nella loro risoluzione. Più tardi, per gli stessi problemi fu usata la trasformata di Laplace, che insieme alle funzioni generatrici fu applicata nello studio di diversi sistemi.

Con Kendall [7] iniziò un approccio probabilistico all'analisi, introducendo processi stocastici nella rappresentazione e nell'analisi dei sistemi di code, sostituendoli ad equazioni di difficile interpretazione. Egli introdusse un metodo per analizzare code non Markoviane, la cui importanza tecnica venne, in seguito, integrata da Cox [8].

Fino agli anni '60 si riscontra e si evidenzia un approccio allo studio dei sistemi di code piuttosto analitico. I teorici, per lo più matematici, sviluppavano processi alquanto complessi e, il più delle volte, coloro che si trovavano a doverli applicare a situazioni reali, riscontravano non poche difficoltà. A fronte di tali problematiche, in quel periodo la teoria delle code, da molti, era ritenuta fosse morta, chiusa, una materia da cui non si potesse trarre altro di interessante [3] [9]. Bhat, nel suo articolo "Sixty years of

Queueing Theory” (1969), si contrappone a questa affermazione, sostenendo che, se ci fosse stata una forte integrazione tra area teorica e area operativa, si sarebbero ottenuti risultati rilevanti e supportava tale previsione con lavori che si stavano iniziando ad effettuare, già in quel periodo, sui problemi di ottimizzazione e di simulazione.

La previsione di Bhat negli anni è stata più che confermata dagli sviluppi che si sono, poi, avuti. Ne rappresentano un esempio quelli nell’ambito delle reti di code, l’ideazione dei modelli di fluidi e di modelli specializzati, utilizzati, prevalentemente, nell’area informatica e di telecomunicazione [9].

Il primo articolo sulle reti di code fu scritto da J.Jackson [10], che propose delle tecniche utili per analizzare processi di traffico nei computer e nelle reti di computer. Ma ben più complesse reti di code furono osservate a partire dagli inizi degli anni ’70 da diversi autori, tra cui, Baskett [11], Kelly [12], Disney e Kiessler [13] e Dai [14].

3.3 sistema di servizio

Da un punto di vista fisico un sistema di servizio è definito come un sistema non vuoto di **servitori** (server), capaci di fornire un servizio alla volta, e da un insieme non vuoto di **aree di attesa** (buffer) capaci di accogliere i **clienti** in arrivo, che non possono essere serviti immediatamente.

Un sistema di servizio è caratterizzato da:

- arrivo *casuale* dei clienti
- tempi necessari per espletare un servizio *non nulli*

Di conseguenza, è molto frequente che un servente non abbia la possibilità di soddisfare immediatamente le richieste di servizio, con conseguente generazione di una fila o coda di clienti in attesa di essere serviti. Infatti, gli utenti che arrivano al sistema (*clienti in arrivo*), in genere, attendono in fila se tutti i serventi (*centro di servizio*) sono occupati, poi vengono serviti ed, infine, lasciano il sistema (*clienti in partenza*). (Fig. 2)

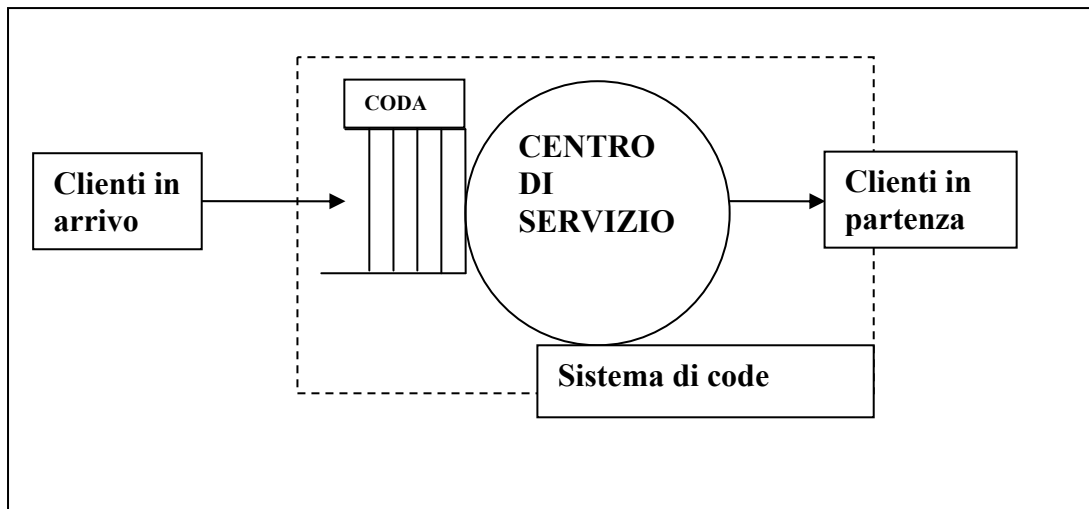


Figura 3.2. Sistema di servizio

Il dimensionamento di un sistema di servizio, semplice o complesso che sia, è realizzato sulla base di alcune grandezze fondamentali, quali, lunghezza media della coda, numero medio di utenti presenti nel sistema, durata media del tempo passato in coda.

Sistemi con tipologie di sistemi semplici possono essere analizzati analiticamente ma, quando il sistema diventa complesso, come nel caso in cui più code sono presenti nel sistema e il servizio richiede operazioni diverse, effettuate da serventi diversi e che devono essere, necessariamente, svolte affinché lo stesso servizio sia completamente erogato, allora, lo studio analitico diventa molto oneroso e l'unica soluzione per stimare le misure di prestazione risiede nell'utilizzo di modelli simulativi.

3.3.1 Strutturazione di un sistema di servizio

Gli elementi che permettono di definire completamente un sistema di servizio e, quindi il fenomeno d'attesa in coda, sono :

- popolazione di utenti
- processo di arrivo
- numero di serventi
- processo di servizio
- capacità del servizio
- disciplina della coda

La popolazione è l'insieme dei potenziali clienti, ovvero l'insieme da cui hanno origine i clienti in arrivo al sistema e a cui ritornano dopo essere stati serviti.

La principale caratteristica della popolazione è la sua *dimensione*, che rappresenta il numero totale dei distinti potenziali clienti che richiedono un servizio.

Se il numero di clienti costituenti la popolazione è influenzato da quelli già presenti all'interno del sistema di code, allora la capacità della popolazione risulta essere finita. Viceversa, quando il numero di clienti in coda non influenza, significativamente, il tasso con il quale la popolazione genera nuovi clienti, la capacità della popolazione è considerata infinita.

Nel primo caso le modalità di arrivo dei clienti dipendono dal numero di quelli correntemente presente nel sistema. Essendo questo il caso più complesso, si assumerà che la dimensione della popolazione sia infinita anche quando è finita, purché sia sufficientemente grande.

I clienti che costituiscono una popolazione, inoltre, sono considerati tra loro indistinguibili. Di conseguenza, si suppone che essi provengano da diverse popolazioni ogni volta presentino caratteristiche diverse, ad esempio livello di priorità o tipologia di servizio richiesto.

Il processo di arrivo descrive il modo secondo il quale i clienti si presentano a richiedere il servizio. Esso è definito in termini di *intertempo di arrivo*, cioè dell'intervallo di tempo intercorrente tra due arrivi successivi. Questo processo può essere di tipo deterministico ma, in genere, è descritto da una variabile aleatoria, indicata con t^a , di cui si suppone nota la distribuzione di probabilità.

Il processo di servizio, invece, descrive il modo secondo il quale ciascun servente eroga il servizio. Esso è definito in termini di *tempo di servizio*, ovvero del tempo necessario ad un servente per realizzare il servizio. Come nel caso dei processi di arrivo, anche questo processo può essere di tipo deterministico, ma nella maggior parte dei casi è descritto da una variabile aleatoria, che indichiamo con t^s , di cui si suppone nota la distribuzione di probabilità. Nel caso in cui il sistema sia composto da più server si assume, in genere che essi siano caratterizzati dallo stesso processo di servizio, ovvero che il tempo di servizio abbia la stessa distribuzione di probabilità per ogni server.

Al fine di ottenere modelli analiticamente trattabili, si assume che sia il processo di arrivo che quello di servizio siano **stazionari**, ovvero che le loro proprietà statistiche siano costanti nel tempo. Precisiamo che, però, tale assunzione può risultare limitativa in certi ambiti, in quanto l'esperienza comune ci insegna, ad esempio, che il processo di arrivo dei clienti in una banca o in un ufficio postale, varia durante le ore della giornata.

Numero di serventi. È fondamentale definire il numero di serventi s presenti nel sistema di servizio al momento della sua analisi. Se è presente più di un servente è indispensabile, inoltre, distinguere se essi lavorano in “serie”, se il servizio richiede più operazioni per essere espletato o in “parallelo”, se necessita di un'unica operazione.

La capacità del servizio corrisponde al numero massimo di utenti che possono essere presenti, contemporaneamente, nel sistema, comprendendo sia gli utenti in coda che quelli che stanno usufruendo del servizio. I clienti, che arrivano dopo che sia saturata questa capacità, sono respinti. Ha capacità di servizio limitata, ad esempio, un centralino telefonico di un call-center che può tenere in attesa solo un numero finito di chiamate. In assenza di centralino, la dimensione della coda è addirittura zero, di conseguenza una chiamata o è servita o è immediatamente respinta.

La disciplina della coda descrive le modalità in base alle quali i clienti vengono “selezionati” dalla coda per essere serviti. Specifica, quindi, quale sarà il prossimo cliente servito tra quelli presenti in coda.

Le discipline di servizio più comuni, sia perché molto vicine alla realtà, sia perché sono matematicamente trattabili, sono:

- **FIFO (“first-in first-out”)**: i clienti vengono serviti nell’ordine in cui arrivano.
- **LIFO (“last-in first-out”)**: che corrisponde a servire per primo l’ultimo cliente arrivato.
- **SIRO (“service in random order”)**: consiste nel servire i clienti in un ordine casuale.
- **SERVIZIO BASATO SU CLASSI DI PRIORITÀ**: è il caso in cui gli utenti vengono raggruppati per classi di priorità in modo tale da servire per primi i clienti appartenenti a quella con priorità più alta; l’esempio tipico, in tal senso, è quello di un pronto soccorso dove sussiste la necessità di intervenire prima sui pazienti più gravi.

Si può, a questo punto, dettagliare un processo di erogazione del servizio e, dunque di formazione di file o code di attesa come segue: ogni utente o cliente, proveniente da una popolazione, accede al servizio in un certo istante e, nel caso in cui i server siano tutti occupati, si unisce alla coda; attende un certo tempo prima di essere servito, terminato il quale, viene selezionato, tra gli utenti presenti nel buffer, secondo la disciplina di servizio vigente. Dopo che il servizio richiesto venga effettuato da uno (o più) server, il cliente lascia il sistema. Se indichiamo con t^w il **tempo complessivo di attesa nel sistema** e con t^q la variabile aleatoria indicante il **tempo di attesa in coda**, risulta che:

$$t^w = t^q + t^s \quad (1.1)$$

Nell’analisi di un sistema di servizio vengono effettuate una serie di assunzioni, come semplificazioni. Si considereranno soddisfatte quelle riguardanti gli **intertempi di arrivo** t^a e i **tempi di servizio** t^s , ritenuti essere **indipendenti e identicamente distribuiti**. Altra assunzione, che si riterrà sempre valida, è relativa ai server che si suppone operino “in parallelo”. In altre situazioni, però, non è da escludere il caso in cui gli stessi server lavorino “in serie” prima di poter completare il servizio. (**Fig. 3 e 4**) [19]

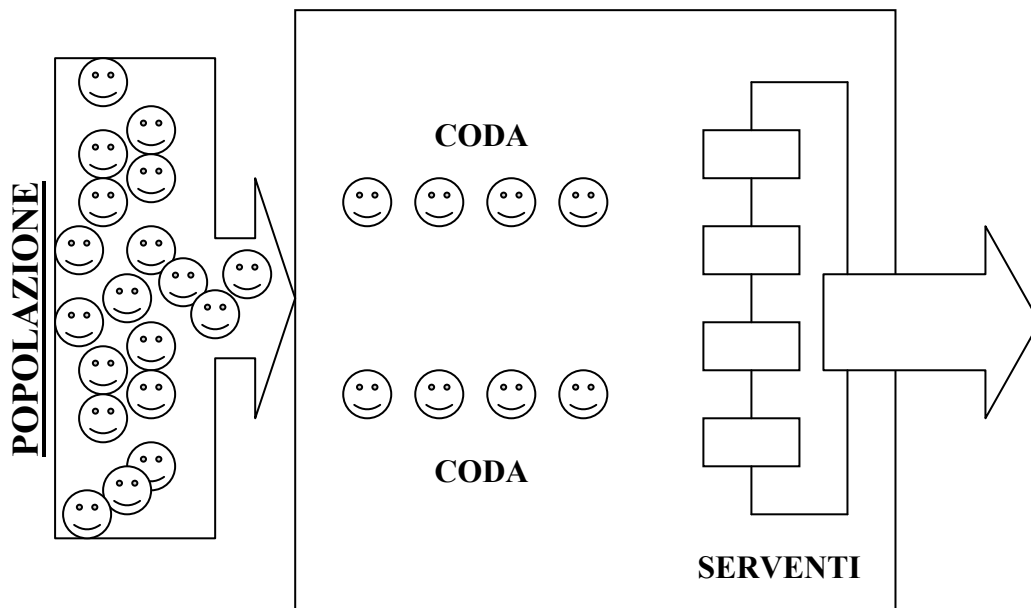


Figura 3.3. Schema di servizio con due code e serventi in parallelo

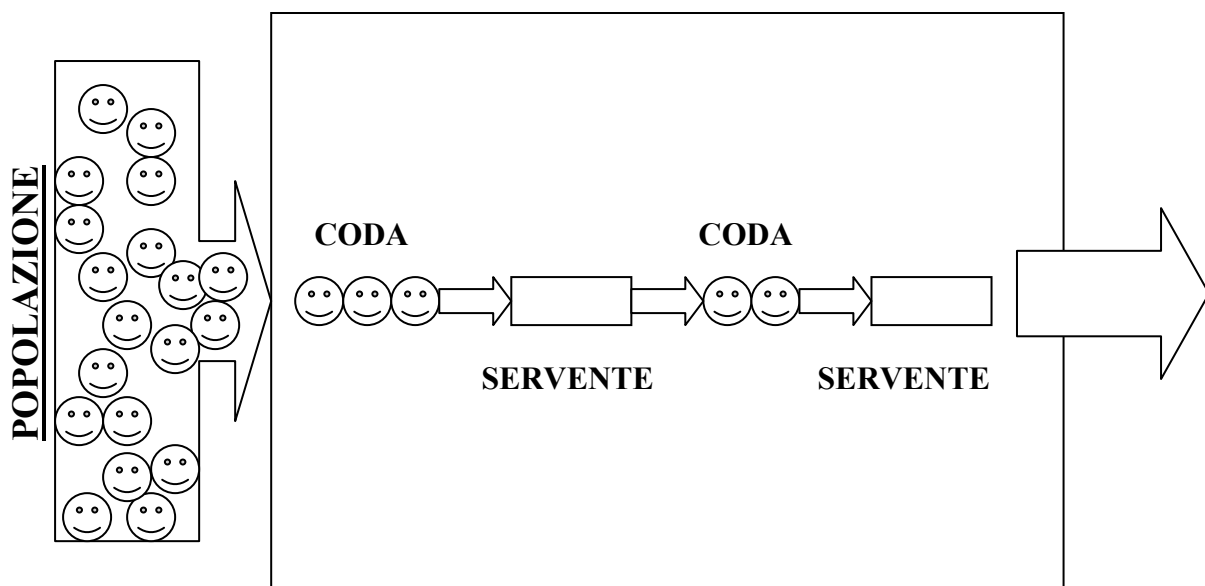


Figura 3.4. Schema di servizio con serventi in serie

3.3.2 Notazione di Kendall

Un modo sintetico per descrivere un sistema di servizio consiste nell'utilizzare la seguente notazione, nota come notazione di Kendall:

$$A / B / s / K / p / Z$$

dove le lettere indicano rispettivamente:

A = distribuzione di probabilità degli intertempi di arrivo

B = distribuzione di probabilità dei tempi di servizio

s = numero di serventi attivi

K = capacità del sistema (default: infinita)

p = dimensione della popolazione (default: infinita)

Z = disciplina di servizio (default: FIFO)

e s, K e p sono numeri interi non negativi.

Se le componenti K, p, e Z non sono specificate, si assumono i valori di default indicati. Le componenti A e B relative alle distribuzioni di probabilità dei processi di arrivo e di servizio sono sostituite dalle seguenti lettere a seconda dei casi:

M: per indicare che sia la distribuzione dei tempi di arrivo che quella dei tempi di servizio è di tipo esponenziale (Markoviano o “memoryless”) (vedi par. 1.5)

D: per indicare la distribuzione costante (degenere) o tempi deterministici.

E_k: per indicare la distribuzione di Erlang di ordine k (vedi par. 1.5)

G: per indicare una distribuzione una distribuzione generica che nel caso degli intertempi di arrivo può essere sostituita dalla sigla **GI** che indica una distribuzione generica di *eventi indipendenti*. [21]

3.3.3 Misure di prestazione e analisi stazionaria

Gli obiettivi che interessano i protagonisti coinvolti in un sistema di servizio, come specificato nei paragrafi precedenti, sono spesso contrastanti tra di loro. I clienti ritengono fondamentale la riduzione dei loro tempi di attesa, mentre, il gestore del servizio deve dimensionare il proprio sistema minimizzando i costi di servizio.

Nel caso in cui il sistema non fosse dimensionato in modo adeguato e, cioè il numero di serventi fosse inferiore al valore ottimale, i tempi di attesa per i clienti possono risultare lunghi. Fenomeni frequenti che possono presentarsi, infatti, sono il **balking**, ossia la rinuncia da parte dell'utente ad entrare nel sistema, e il **reneging**, ossia l'abbandono del cliente dopo essersi unito alla coda. In questi casi, il fornitore del servizio sosterebbe un costo indiretto dovuto all'eventuale mancato guadagno.

In questo contesto, dunque, la teoria delle code è finalizzata alla determinazione delle distribuzioni di probabilità di alcune variabili aleatorie che interessano le principali misure di prestazione di un sistema di servizio, e che, sotto determinate ipotesi, sono facilmente calcolabili.

Note queste distribuzioni è possibile poi risalire ai costi che ne conseguono, ovvero al costo del personale e delle attrezzature necessarie per erogare il servizio da parte del gestore, o il costo, in termini di tempo passato in attesa, da parte dei clienti.

Prima di analizzare nel dettaglio le misure di prestazione che interessano un sistema di servizio, è indispensabile definire ancora qualche aspetto legato alla struttura del sistema stesso. In particolare, risulta necessario specificare:

- **lo stato di una coda.** In genere, lo stato di un sistema dinamico in un dato istante temporale rappresenta l'insieme informativo minimo che permette di conoscere l'evoluzione futura del sistema stesso, una volta note le realizzazioni dei fenomeni stocastici cui è soggetto.

Nel caso di un sistema di code lo stato rappresenta il numero complessivo di clienti presenti nel sistema stesso ed è, quindi, dato dalla somma del numero di clienti che sono in coda e il numero di serventi attivi.

Lo stato del sistema lo indicheremo con la lettera n .

- **lunghezza di una coda.** Indicata con n^q , rappresenta il numero di clienti presenti in coda o in attesa di essere serviti. La lunghezza di una coda dipende dal numero di serventi s e dallo stato del sistema n . In particolare:

$$n^q = \begin{cases} 0 & \text{se } n \leq s \\ n-s & \text{se } n > s \end{cases} \quad (3.2)$$

- **frequenza media di arrivi λ .** Denota il numero medio di clienti che arriva nel sistema nell'unità di tempo.

- **velocità di servizio μ .** Rappresenta il numero medio di clienti serviti nell'unità di tempo.

Indicando con $E(t^a)$ e con $E(t^s)$ i valori attesi rispettivamente delle variabili aleatorie t^a e t^s , risulta che:

$$\lambda = 1 / E(t^a) \quad (1.3)$$

$$\mu = 1 / E(t^s) \quad (1.4)$$

- **coefficiente di utilizzazione dei serventi ρ .** E' il rapporto tra la frequenza media di arrivo λ e la velocità di servizio μ moltiplicata per il numero di serventi :

$$\rho = \lambda / s\mu \quad (1.5)$$

Tale rapporto rappresenta, quindi, la frazione di tempo durante la quale tutti i serventi risultano occupati; il suo valore, pertanto, denota la capacità di servizio utilizzata in media dagli utenti che arrivano.

A partire da tali definizioni, l'analisi di un sistema di servizio prosegue considerando una serie di grandezze fondamentali assunte come misure di prestazione.

Tali grandezze verranno considerate assumendo che il sistema abbia raggiunto una **condizione stazionaria** o **di equilibrio** (stady-state). Tale condizione denota la situazione in cui il sistema è operativo da un tempo sufficientemente grande da risultare indipendente dalle condizioni iniziali e dal tempo che è trascorso dall'attivazione del sistema stesso, come invece avviene in **condizioni transitorie**. Quest'ultime risultano più complesse da un punto di vista analitico, per cui, in genere, poco trattate.

Risulta evidente che non è possibile raggiungere una condizione di equilibrio se risulta che $\rho \geq 1$; in tal caso, infatti, la frequenza media di arrivo supererebbe la capacità del servizio e lo stato crescerebbe indefinitamente.

Se invece è raggiunto l'equilibrio, la distribuzione di probabilità dello stato rimane costante nel tempo.

L'analisi di un sistema di servizio in condizioni stazionarie prevede che vengano considerate i valori attesi delle grandezze definite in precedenza, ovvero:

- p_n : probabilità che lo stato del sistema sia n
- N : valore atteso dello stato del sistema (n)
- T : valore atteso del tempo passato nel sistema (t^a)
- N^q : valore atteso del numero di clienti in coda (n^q)
- T^q : valore atteso del tempo passato in coda (t^q).

Essendo lo stato del sistema n e la lunghezza della coda n^q due variabili aleatorie discrete, dalla definizione di valore atteso [19] risulta che:

$$N = E(n) = \sum_{n=0}^{\infty} n p_n \quad (1.6) \quad N^q = E(n^q) = \sum_{n=s+1}^{\infty} (n-s) p_n \quad (1.7)$$

Essendo invece t^w e t^q due variabili aleatorie continue, indicando con $f_{t^w}(t)$ e con $f_{t^q}(t)$ le rispettive densità di probabilità, si ha:

$$T = E(t^w) = \int_0^{\infty} t f_{t^w}(t) dt \quad (1.8) \quad T^q = E(t^q) = \int_0^{\infty} t f_{t^q}(t) dt \quad (1.9)$$

Le variabili aleatorie che descrivono le grandezze \mathbf{n} ed \mathbf{n}^q , così come la stessa probabilità \mathbf{p}_n che lo stato sia n , costituiscono in realtà dei *processi stocastici* dal momento che dipendono dal parametro t che indica il tempo.

In tal senso, per ogni istante di tempo $t \geq 0$ è possibile definire le variabili $\mathbf{N}(t)$ e $\mathbf{N}^q(t)$ che rappresentano rispettivamente lo stato del sistema e lo stato della coda all'istante t . Analogamente anche \mathbf{p}_n dipende dal tempo, per cui si definisce la probabilità $\mathbf{p}_n(t)$ che lo stato del sistema all'istante t sia n .

In particolare risulta che il valore atteso:

$$\mathbf{E}[\mathbf{N}(t)] = \sum_{n=0}^{\infty} n \mathbf{p}_n(t) \quad (1.10)$$

e la probabilità $\mathbf{p}_n(t)$ dipendono sia dal tempo t che dalle distribuzioni di probabilità iniziali $\{p_0(0), p_1(0), \dots\}$. [19]

Tale dipendenza si perde nel momento in cui il sistema, operativo da un tempo sufficientemente lungo, raggiunge una condizione di equilibrio; in tal caso, indipendentemente dalla stato iniziale risulta che:

$$\lim_{t \rightarrow \infty} \mathbf{p}_n(t) = \mathbf{p}_n \quad \text{per } n = 0, 1, \dots \quad (1.11)$$

$$\lim_{t \rightarrow \infty} \mathbf{E}[\mathbf{N}(t)] = \mathbf{N} \quad \text{per } n = 0, 1, \dots \quad (1.12)$$

Allo stesso modo si può definire il valore di $\mathbf{E}[\mathbf{N}^q(t)]$, ovvero:

$$\mathbf{E}[\mathbf{N}^q(t)] = \sum_{n=s+1}^{\infty} (n-s) \mathbf{p}_n(t) \quad (1.13)$$

che nel caso il sistema raggiunga l'equilibrio, ossia per t che tende all'infinito, risulta:

$$\lim_{t \rightarrow \infty} \mathbf{E}[\mathbf{N}^q(t)] = \mathbf{N}^q \quad (1.14)$$

Per quanto riguarda il tempo di attesa nel sistema \mathbf{t}^w e il tempo passato in coda \mathbf{t}^q , anche in questo caso si individua una famiglia di variabili aleatorie \mathbf{t}_i^w e \mathbf{t}_i^q , rispettivamente per \mathbf{t}^w e \mathbf{t}^q ; in particolare:

\mathbf{t}_i^w = tempo passato nel sistema dall' i -esimo cliente

\mathbf{t}_i^q = tempo passato nella coda dall' i -esimo cliente.

Al tendere di i all'infinito, il valore atteso $\mathbf{E}(\mathbf{t}_i^w)$ avrà un valore di equilibrio:

$$\lim_{i \rightarrow \infty} E(t_i^w) = T \quad (1.15)$$

e, analogamente, il valore atteso del tempo passato in coda dall'i-esimo cliente, $E(t_i^q)$, all'equilibrio tende al valore T^q , ovvero:

$$\lim_{i \rightarrow \infty} E(t_i^q) = T^q. \quad (1.16)$$

Si riassumono le principali misure di prestazione e i loro valori attesi in condizioni di stazionarietà nella **Tab.3**.

Misure di prestazione	Notazione	Valori stazionari
Stato del sistema al tempo t	$N(t)$	$\lim_{t \rightarrow \infty} E[N(t)] = N$
Lunghezza della coda al tempo t	$N^q(t)$	$\lim_{t \rightarrow \infty} E[N^q(t)] = N^q$
Tempo passato nel sistema dall'i-esimo cliente	t_i^w	$\lim_{t \rightarrow \infty} E(t_i^w) = T$
Tempo passato in coda dall'i-esimo cliente	t_i^q	$\lim_{t \rightarrow \infty} E(t_i^q) = T^q$

Tabella 3.1. Misure di prestazione di un sistema di servizio

3.3.4 Teorema di Little

Tra le misure di prestazione che caratterizzano un sistema di servizio esistono importanti relazioni, valide nell'ipotesi in cui il sistema di code abbia raggiunto la condizione di equilibrio.

In particolare, la più importante relazione che si stabilisce tra i parametri N , T , N^q e T^q , precedentemente definiti, è nota come **teorema (o legge) di Little**, secondo cui, in un sistema di code in *condizioni stazionarie* valgono le seguenti relazioni:

$$N = \lambda T \quad (1.17)$$

$$N^q = \lambda T^q \quad (1.18)$$

dove N e N^q indicano, rispettivamente, il numero medio di clienti nel sistema e il numero medio dei clienti in coda, mentre T e T^q il tempo medio trascorso nel sistema stesso e in coda e λ la frequenza media di arrivi.

E' importante sottolineare che la frequenza λ deve essere quella *effettiva*, ovvero quella degli ingressi effettivi nel sistema. Spesso, infatti, per alcune tipologie di servizi la frequenza media degli arrivi potrebbe non coincidere con la frequenza degli ingressi effettivi; si pensi, ad esempio, ai sistemi a capacità limitata con possibilità di rinuncia al servizio.

La legge di Little afferma, quindi, che il numero medio di utenti in un sistema a coda in equilibrio è pari al tasso medio di arrivo degli utenti al sistema moltiplicato per il tempo medio speso nel sistema.

Un'altra importante relazione tra il tempo medio passato nel sistema da un cliente e il tempo medio passato in coda è la seguente:

$$T = T^q + \frac{1}{\mu} \quad (1.19)$$

che può essere facilmente verificata considerando il tempo complessivo passato da un cliente nel sistema, dato dall'espressione (1.1), e passando ai valori attesi delle variabili t^w , t^q e t^s . [43]

3.3.5 Proprietà PASTA

Per sistemi di code con **arrivi Poissoniani** (vedi par. 1.5.2), cioè per i sistemi $M/\cdot/\cdot$, è importante menzionare una proprietà di una certa rilevanza, detta **proprietà PASTA (Poisson Arrivals See Times Average)**. Secondo tale proprietà, i clienti in arrivo al sistema con una distribuzione degli intertempi di arrivo di tipo esponenziale, trovano in media la stessa situazione che un osservatore esterno ad esso troverebbe in un istante generico t . Si consideri che $a_k(t)$ sia la probabilità che un utente che arriva al tempo t trovi il sistema nello stato k . Essa, in genere, è diversa dalla probabilità che lo stato del sistema in un generico istante t sia pari a k , ovvero:

$$\mathbf{a}_k(\mathbf{t}) \neq \mathbf{p}_k(\mathbf{t}). \quad (1.20)$$

Ma, nel caso in cui gli arrivi siano Poissoniani, le due probabilità risultano essere uguali:

$$\mathbf{a}_k(\mathbf{t}) = \mathbf{p}_k(\mathbf{t}). \quad (1.21)$$

Tale uguaglianza è l'essenza della *proprietà PASTA*. [19]

All'equilibrio, quando il sistema è operativo da un tempo sufficientemente lungo, si perde la dipendenza dalle condizioni iniziali e dal tempo t e risulterà:

$$\lim_{t \rightarrow \infty} \mathbf{p}_k(\mathbf{t}) = \mathbf{p}_k \quad (1.22)$$

e

$$\lim_{t \rightarrow \infty} \mathbf{a}_k(\mathbf{t}) = \mathbf{a}_k \quad (1.23)$$

che si interpretano, rispettivamente, come la frazione di tempo che il sistema è nello stato k e la frazione di arrivi che trovano k utenti nel sistema. Quindi, la *proprietà PASTA* allo *stato stazionario* sarà

$$\mathbf{a}_k = \mathbf{p}_k. \quad (1.24)$$

Si consideri, ora, la probabilità che un utente allo stato t lasci il sistema nello stato k e la si indica con $\mathbf{d}_k(\mathbf{t})$. In condizioni di equilibrio si avrà

$$\lim_{t \rightarrow \infty} \mathbf{d}_k(\mathbf{t}) = \mathbf{d}_k \quad (1.25)$$

che rappresenta la frazione di clienti che lascia nel sistema k clienti quando vi esce.

In genere, per un qualsiasi sistema a coda singola in cui i clienti entrano uno alla volta e vengono serviti uno alla volta si ha che

$$\mathbf{a}_k = \mathbf{d}_k. \quad (1.26)$$

Nel caso in cui gli arrivi sono Poissoniani e vale la *proprietà PASTA*, si avrà che, in condizioni di stazionarietà, sia un cliente che arriva, sia uno che parte da un sistema trova un sistema che è statisticamente equivalente a quello che un osservatore esterno vede in un generico istante temporale [19], ovvero

$$\mathbf{a}_k = \mathbf{d}_k = \mathbf{p}_k. \quad (1.27)$$

3.4 Processi Stocastici Di Arrivo E Di Servizio

Per modellare un sistema di servizio, come già detto, occorre rappresentarlo attraverso le sue componenti e, in particolare, è importante poter individuare i processi di arrivo e di servizio che lo caratterizzano.

Prima che gli utenti inizino la loro attesa e/o vengano serviti, devono entrare nel sistema. Essi possono arrivare uniformemente o in modo imprevedibile; possono arrivare uno alla volta o in gruppo. In ogni caso, la modalità con cui giungono al sistema caratterizza il loro *processo di arrivo*.

Un processo di arrivo particolare, che trova ampia applicazione nei modelli di code, è il **processo di arrivo Markoviano**, con il quale si rappresenta una situazione in cui i clienti arrivano uno alla volta e dove gli intertempi di arrivo sono variabili aleatorie di tipo esponenziale.

Una volta che i clienti sono entrati nel sistema, devono essere serviti. Il *processo di servizio* rappresenta la modalità con cui ciò avviene, se quindi vi sono uno o più serventi, se il servizio richiede più operazioni per essere completato, con quale modalità vengono selezionati i clienti da servire, se i tempi di servizio sono troppo lunghi. Anche in questo caso, il processo maggiormente utilizzato è **Markoviano**, in cui i clienti sono serviti uno alla volta, seguendo una disciplina di tipo FIFO, e tempi di servizio di tipo esponenziale.

Il processo di servizio è alimentato da quello di arrivo che è indipendente dal primo, ma lo condiziona. Un cliente, infatti, può essere servito solo se è già entrato nel sistema. In assenza di utenti, dunque, il server rimane inattivo e, quindi, non può avvantaggiarsi in vista di impegni futuri. In altri termini, un server non può servire in anticipo clienti non ancora arrivati, e quindi, non può esistere una coda negativa.

3.4.1 Il ruolo della distribuzione esponenziale

Si è detto che i processi di arrivo e di servizio più comuni sono di tipo Markoviano, con intertempi di arrivo e di servizio variabili aleatorie di tipo esponenziale. Ciò è dovuto alle sue caratteristiche, ma in

particolare, per la **proprietà di assenza di memoria** di cui è l'unica, tra le variabili casuali di tipo continuo, a goderne.

Si ricordi che una v.a. continua T segue una **distribuzione** di tipo **esponenziale** con parametro α se la funzione di distribuzione (Cdf) e quella densità di probabilità (Pdf) [20] sono rispettivamente (**Fig. 5**):

$$F_T(t) = P(T \leq t) = \begin{cases} 1 - e^{-\alpha t} & \text{per } t \geq 0 \\ 0 & \text{per } t < 0 \end{cases} \quad (1.28) \quad \text{e} \quad f_T(t) = \begin{cases} \alpha e^{-\alpha t} & \text{per } t \geq 0 \\ 0 & \text{per } t < 0 \end{cases} \quad (1.29)$$

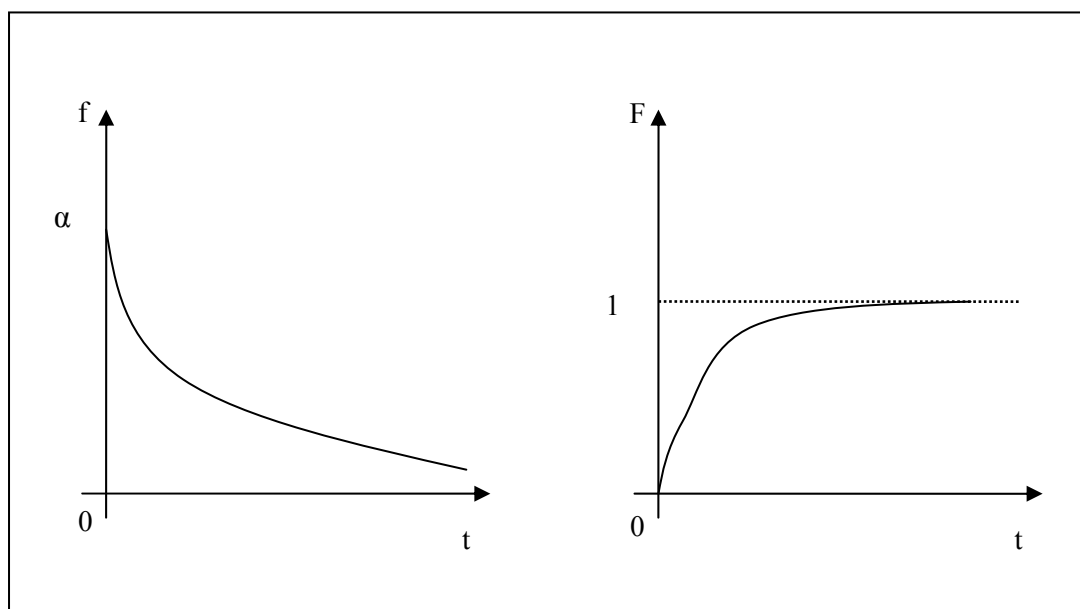


Figura 3.5. Pdf e Cdf di una variabile aleatoria esponenziale

ed è caratterizzata da valore atteso e varianza date da:

$$E(T) = \frac{1}{\alpha} \quad (1.30) \quad \text{e} \quad Var(T) = \frac{1}{\alpha^2} \quad (1.31).$$

La caratteristica principale di questo modello di variabile aleatoria, come già detto, è la proprietà di assenza di memoria. In genere, fissati due valori t_1 ed

$t_2 = t_1 + t$, la probabilità che non si verificano eventi tra t_1 e t_2 , supposto che non se ne siano verificati fino a t_1 , è pari a:

$$\begin{aligned} P\{T > t_2 | T > t_1\} &= \frac{P\{(T > t_2) \cap (T > t_1)\}}{P\{T > t_1\}} = \frac{P\{T > t_2\}}{P\{T > t_1\}} = \frac{1 - F_T(t_2)}{1 - F_T(t_1)} = \\ &= \frac{1 - F_T(t_1 + t)}{1 - F_T(t_1)} = \frac{e^{-\alpha(t_1+t)}}{e^{-\alpha(t_1)}} = e^{-\alpha t}. \end{aligned} \quad [20] \quad (1.32)$$

Utilizzando la Cdf esponenziale notiamo, quindi, che la probabilità suddetta dipende solo dalla distanza t , tra t_1 e t_2 , e non dai loro valori.

Nell'ambito della teoria delle code ciò sta ad indicare che la probabilità del tempo che rimane fino ad un prossimo evento, sia esso un arrivo o il completamento di un servizio, è sempre la stessa, indipendentemente da quanto tempo è trascorso. Tale mancanza di memoria della distribuzione esponenziale fa sì che essa sia utile per modellare gli intertempi di arrivo t^a che non siano correlati, cioè per cui l'arrivo di un cliente non influenzi altri arrivi, e i tempi di servizio t^s che riguardano prestazioni poco omogenee, ad esempio i servizi di pronto soccorso che risultano essere diversi da paziente a paziente. [19]

Altra caratteristica di una variabile aleatoria esponenziale è che, considerate n variabili aleatorie T_1, T_2, \dots, T_n distribuite esponenzialmente con parametri $\alpha_1, \alpha_2, \dots, \alpha_n$ ed indipendenti, cioè tali che la loro distribuzione di probabilità (o di densità) congiunta sia pari al prodotto delle singole funzioni di distribuzione (o di densità) marginali, una variabile casuale U pari al minimo tra le n variabili

$$U = \min \{T_1, T_2, \dots, T_n\}, \quad (1.33)$$

sarà distribuita esponenzialmente con parametro

$$\alpha = \sum_{i=1}^n \alpha_i. \quad (1.34)$$

Tale proprietà è importante per modellare i tempi di servizio in un sistema di code. Si considerino s serventi in parallelo, ognuno dei quali abbia tempi di servizio distribuiti esponenzialmente con lo stesso parametro μ e che solo $r \leq s$ serventi stia attualmente fornendo il servizio e T_i sia il tempo che rimane per il completamento da parte di ognuno ($i = 1, \dots, r$). Per la proprietà sopra esposta, T_i sarà distribuita esponenzialmente con parametro $\alpha_i = \mu$ e U , che è la variabile rappresentante il tempo fino al prossimo completamento di servizio, sarà anch'essa una variabile casuale esponenziale ma con parametro $\alpha = r\mu$.

Quindi, studiare un sistema di code con più server posti in parallelo, ciascuno con tempo di servizio esponenziale di parametro μ , equivale ad analizzare un sistema a singolo server con tempo di servizio distribuito esponenzialmente con parametro μ .

Considerata S_n la somma delle variabili aleatorie esponenziali identicamente distribuite con parametro μ , T_1, T_2, \dots, T_n , ovvero

$$S_n = T_1 + T_2 + \dots + T_n \quad (1.35)$$

essa avrà la seguente densità di probabilità

$$f_{S_n}(t) = \mu e^{-\mu t} \frac{(\mu t)^{n-1}}{(n-1)!} \quad (1.36)$$

Una distribuzione di probabilità che presenta tale densità di probabilità si chiama **distribuzione di Erlang** di parametri μ e n . Essa si riduce a quella esponenziale se $n = 1$. [21]

3.4.2 Il processo di Poisson

Nell'ambito dei sistemi di code, per modellare i processi di arrivo degli utenti al sistema, viene spesso utilizzato un processo di tipo Poissoniano. Esso è un processo stocastico di conteggio $\{X(t), t \geq 0\}$, ossia che “conta” il numero totale di eventi $X(t)$ accaduti fino al tempo t . Ad esempio, $X(t)$ può rappresentare il numero di clienti entrati in un supermercato fino al tempo t .

In un processo di conteggio valgono le seguenti proprietà:

- $X(t) \geq 0 \quad \forall t \geq 0$
- $X(t)$ ha valori interi
- se $s \leq t$ $X(s) \leq X(t)$
- per $s < t$, $X(t) - X(s)$ è il numero degli eventi che sono accaduti nell'intervallo $[s, t]$.

Il processo di Poisson di parametro λ per essere tale deve presentare anche altre caratteristiche, quali:

- $X(0) = 0$, ossia il conteggio deve partire dal valore zero
- il processo deve presentare incrementi indipendenti, ossia il numero di eventi che accadono in intervalli di tempo disgiunti sono indipendenti

- il numero di eventi $X(s+t)-X(s)$ in ogni intervallo di tempo di ampiezza t , ha distribuzione di Poisson di parametro λt e valore atteso $E\{X(t) = n\} = \lambda t$, cioè $\forall s, t \geq 0$:

$$P\{X(s+t) - X(s) = n\} = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \quad n = 0, 1, \dots \quad (1.37)$$

dove λ esprime il numero medio di eventi nell'unità di tempo e, quindi, rappresenta una frequenza media (**Fig. 6**).

Quest'ultimo punto equivale a dire che il processo ha incrementi stazionari, ovvero la distribuzione del numero di eventi che accadono in un intervallo di tempo dipende solo dall'ampiezza di tale intervallo.

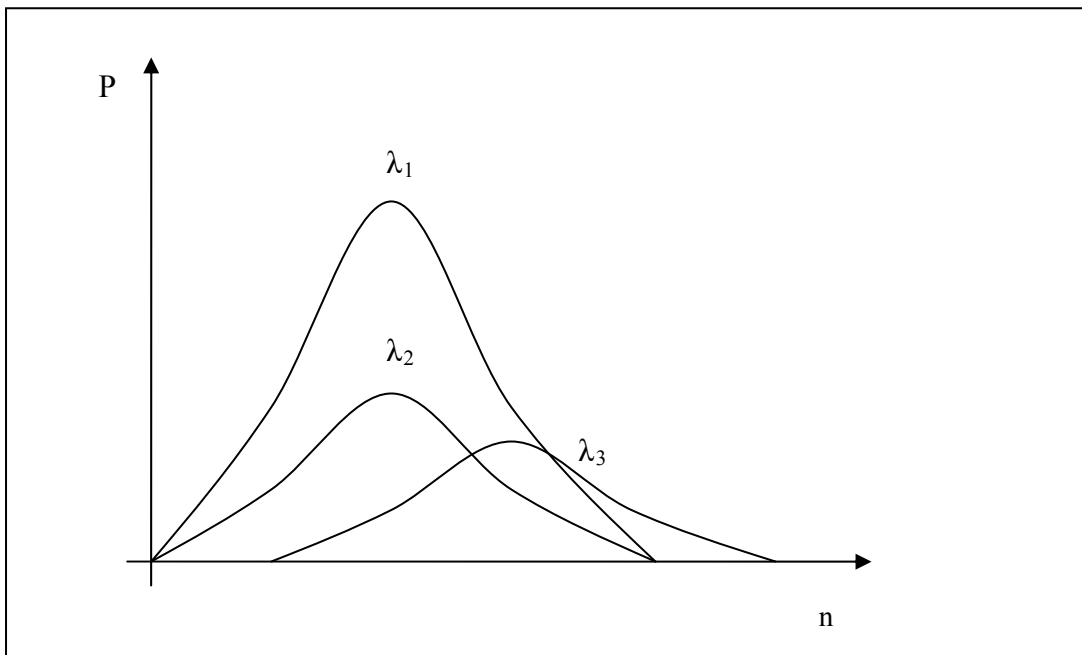


Figura 3.6. Distribuzioni di Poisson con diversi valori λ_i

Nella teoria delle code con tale processo viene rappresentato l'aspetto probabilistico degli arrivi dei clienti al sistema quando gli intertempi di arrivo t^a sono distribuiti esponenzialmente con parametro λ , che rappresenta la frequenza media degli **arrivi**, detti, **poissoniani**. $\{X(t), t \geq 0\}$, ossia il processo di Poisson come definito prima, rappresenterà il numero di clienti in arrivo fino ad un istante t . [21]

Può essere effettuato lo stesso discorso per gli eventi che evidenziano il completamento di un servizio, in cui il tempo di servizio t^s sia distribuito esponenzialmente con parametro μ , che rappresenta la velocità di servizio. In questo caso, il processo stocastico di Poisson rappresenterà il numero di servizi portati al termine fino ad un tempo t .

Mediante l'utilizzo del processo di Poisson è possibile modellare arrivi casuali nel tempo che permettono una rappresentazione alquanto realistica rispetto al caso in cui utilizzassimo altre distribuzioni. Da qui la diffusione ampia nella Teoria delle code degli arrivi poissoniani. In **figura 7** si mostra un esempio di realizzazione di un processo di Poisson e un processo di arrivo con intertempi di arrivo rappresentati da una Erlang-10 (vedi par. 1.5.1). La figura mostra che gli arrivi secondo Erlang sono più egualmente distribuiti rispetto a quelli poissoniani.

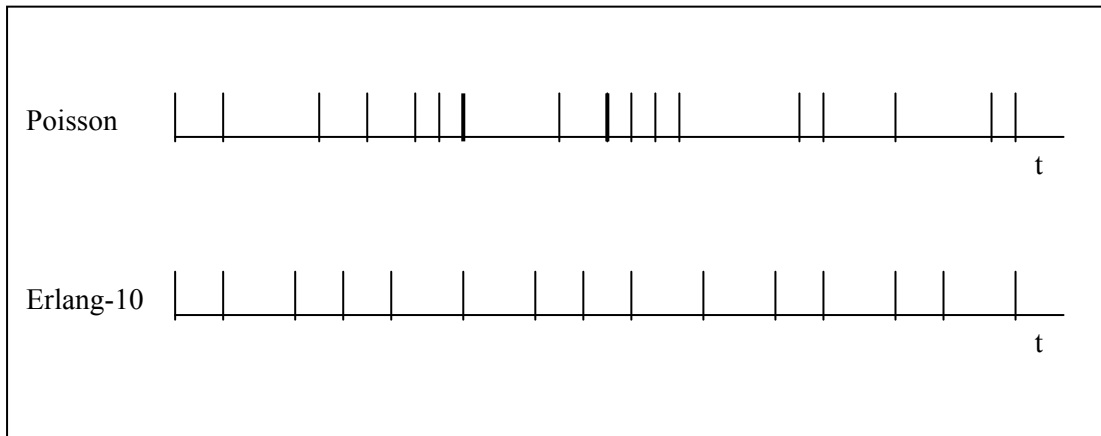


Figura 3.7. Esempio di realizzazione di arrivi Poissoniani e di Erlang-10

Ora si supponga che ogni evento che accade possa essere classificato di tipo A con probabilità p e di tipo B con probabilità $1-p$ e si indichi con $X_A(t)$ il numero degli eventi di tipo A e con $X_B(t)$ il numero di eventi di tipo B fino all'istante t tali che

$$X(t) = X_A(t) + X_B(t) \quad (1.38)$$

dove $X(t)$ si definisce *processo aggregato* e $X_A(t)$ e $X_B(t)$ *processi disaggregati*. Se $\{X(t), t \geq 0\}$ è un processo di Poisson, allora risulterà che i processi disaggregati di cui è costituito, ovvero

$$\{X_A(t), t \geq 0\} \text{ e } \{X_B(t), t \geq 0\} \quad (1.39)$$

sono anch'essi processi di Poisson rispettivamente di parametro λp e $\lambda(1-p)$ ed indipendenti.

Nella teoria delle code si ha che se un sistema ha r tipologie di clienti in arrivo e, quindi r processi di arrivo disaggregati, essi sono processi poissoniani indipendenti e di parametro λ_i con $i = 1 \dots r$ ed il processo aggregato sarà di Poisson con parametro

$$\lambda = \sum_{i=1}^r \lambda_i . \quad (1.40)$$

Tipico esempio di arrivi disaggregati è in un pronto soccorso dove coloro che arrivano per ricevere assistenza medica avranno priorità diverse ad essere serviti (codice rosso, giallo e verde).

3.5 Caratterizzazione Dei Processi Di Nascita E Morte Nei Sistemi Di Servizio

Il *processo di nascita e morte* è un caso particolare di un processo di Markov tempo-continuo [19], dove gli stati rappresentano l'attuale dimensione della popolazione e dove le transazioni indicano *nascita* e *morte*.

Con tale processo si rappresenta il numero di elementi $N(t)$ di una popolazione che può aumentare, per effetto di una nascita, o diminuire, per effetto di una morte, di un'unità alla volta.

La dinamica di un sistema di servizio prevede che il numero di individui (clienti) presenti nel sistema stesso al tempo t , costituisca un processo stocastico $N(t)$ (dimensione della popolazione) in cui la nascita è rappresentata da un nuovo arrivo e la morte da una partenza dal sistema di un utente dopo aver usufruito del servizio.

Volendo analizzare tale dinamica attraverso i processi di nascita e morte, risulta indispensabile definire la probabilità che l'insieme, costituito da n elementi al tempo t , sia costituito da m elementi nel tempo $t + \Delta t$, ovvero:

$$p_{n,m}(\Delta t) = P(N(t + \Delta t) = m | N(t) = n) \quad (1.41)$$

tale probabilità prende il nome di *probabilità di transizione*. Si consideri $p_{n,m}(\Delta t)$ dipendenti dal valore dello stato assunto in t , e non da quello assunto negli istanti precedenti (*proprietà Markoviana o di assenza di memoria*). Definiti i seguenti parametri:

$\lambda_i \geq 0$ = il tasso di nascita allo stato i -esimo (*coefficiente di natalità*)

$\mu_i \geq 0$ = il tasso di morte allo stato i -esimo (*coefficiente di mortalità*)

si possono rappresentare tramite le probabilità di transizione da uno stato i -esimo a uno j -esimo, gli eventi arrivo e partenza dal sistema di un cliente in un intervallo di tempo $(t, t + \Delta t)$:

$$p_{i,j}(\Delta t) = \begin{cases} \mu_i \Delta t & j = i-1 \\ \lambda_i \Delta t & j = i+1 \\ 1 - (\lambda_i + \mu_i) \Delta t & j = i \\ 0 & \text{altrimenti} \end{cases} \quad (1.42)$$

Tali condizioni possono essere così interpretate: la probabilità che, nell'intervallo $[t, t+\Delta t]$ si verifichi una nascita, non dipende da t , ma è proporzionale all'ampiezza dell'intervallo (ovvero della durata di Δt) secondo un coefficiente che può dipendere dallo stato attuale, ma non da quelli passati.

Analogamente, la probabilità che nell'intervallo $[t, t+\Delta t]$ si verifichi una morte è proporzionale alla durata Δt secondo un coefficiente che può dipendere dallo stato attuale, ma non da quelli passati.

Da un punto di vista grafico i processi di nascita e morte possono essere rappresentati attraverso grafi orientati, in cui ogni nodo è contraddistinto da un indice che denota lo stato del sistema, ed ogni arco, invece indica la probabilità di transizione dallo stato corrente a quello successivo o precedente.

La **figura 8** rappresenta le probabilità suddette tramite grafo.

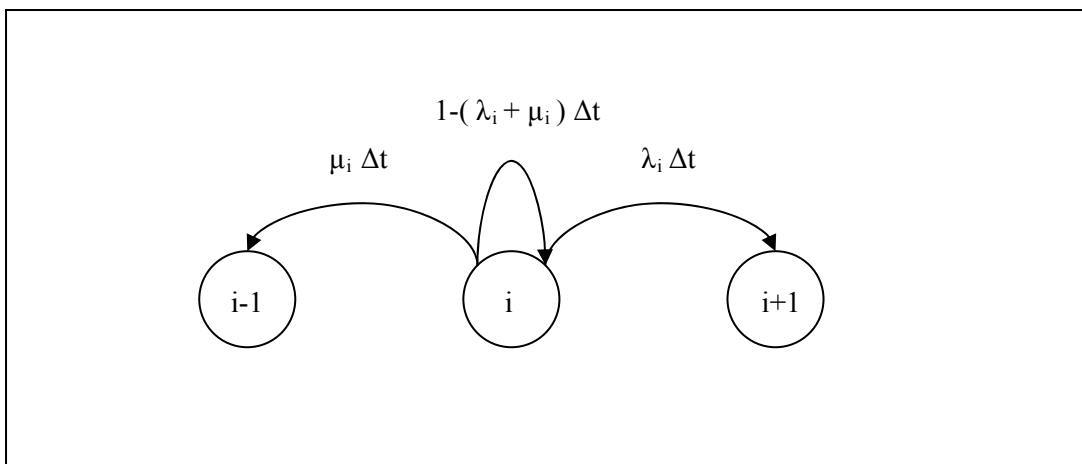


Figura 3.8. Rappresentazione tramite grafo orientato delle probabilità di transizione

Si riporta di seguito il diagramma di transizione di stato di un processo di nascita e morte quando $\Delta t \rightarrow 0$ (**Fig. 9**).

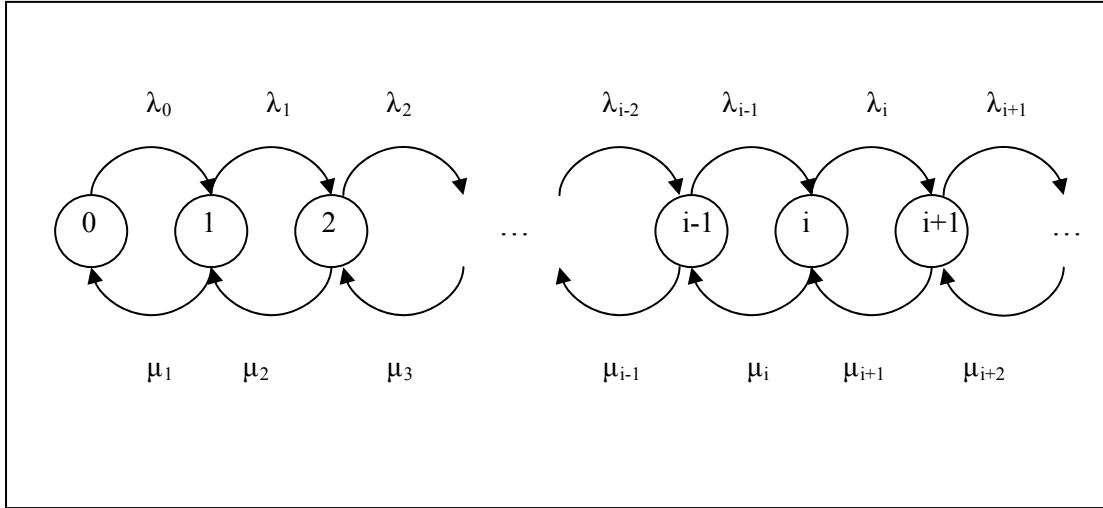


Figura 3.9. Diagramma di transizione di stato di un processo di nascita e morte

Note le probabilità di transizioni $p_{i,j}(\Delta t)$ si arriverà a definire un sistema di equazioni differenziali che nel caso stazionario diventa un sistema di equazioni algebriche di semplice risoluzione, definite **equazioni di Kolmogorov**:

da cui

$$\begin{cases} dp_0(t) / dt = - p_0(t) \lambda_0 + p_1(t) \mu_1 \\ dp_i(t) / dt = - p_i(t) (\lambda_i + \mu_i) + p_{i+1}(t) \mu_{i+1} + p_{i-1}(t) \lambda_{i-1} \end{cases} \quad (1.43)$$

che descrivono la dinamica del processo di nascita e morte, ovvero come evolvono nel tempo le probabilità dei diversi stati del sistema. Il numero di tali equazioni risulta infinito, pertanto la loro risoluzione risulta onerosa e poco pratica. [19]

Ci si limita, dunque, a dire che il set di equazioni (1.43) può essere risolto conoscendo le condizioni iniziali date dalle probabilità $p_i(t)$, $i=0,1,\dots$ e ricordando che $\sum_{i=0}^{\infty} p_i(0) = 1$.

Il modo più conveniente per risolvere il sistema di equazioni di Kolmogorov risiede nel considerare il limite per $t \rightarrow \infty$ ad ambo i membri di ciascuna delle equazioni del sistema sapendo che $\lim_{t \rightarrow \infty} p_i(t) = p_i$ e che per la stazionarietà vale :

$$\lim_{t \rightarrow \infty} \frac{dp_i(t)}{dt} = 0 \quad i=0,1,\dots \quad (1.44)$$

dal momento che a regime le $p_n(t)$ tendono a valori costanti e ricordando che:

$$\sum_{i=0}^{\infty} p_i(t) = 1 \quad (1.45)$$

In tal caso il sistema di equazione differenziali diventa un sistema di equazioni algebriche, ovvero:

$$\begin{cases} p_0 \lambda_0 = p_1 \mu_1 \\ p_i (\lambda_i + \mu_i) = p_{i+1} \mu_{i+1} + p_{i-1} \lambda_{i-1} \end{cases} \quad (1.46)$$

Da queste relazioni, risulta che $p_i(t + \Delta t)$ è calcolata attraverso un bilanciamento dei flussi entranti nello stato i ed uscenti da i stesso. Disegnando un contorno chiuso intorno allo stato zero (**Fig. 10**), si avrà,

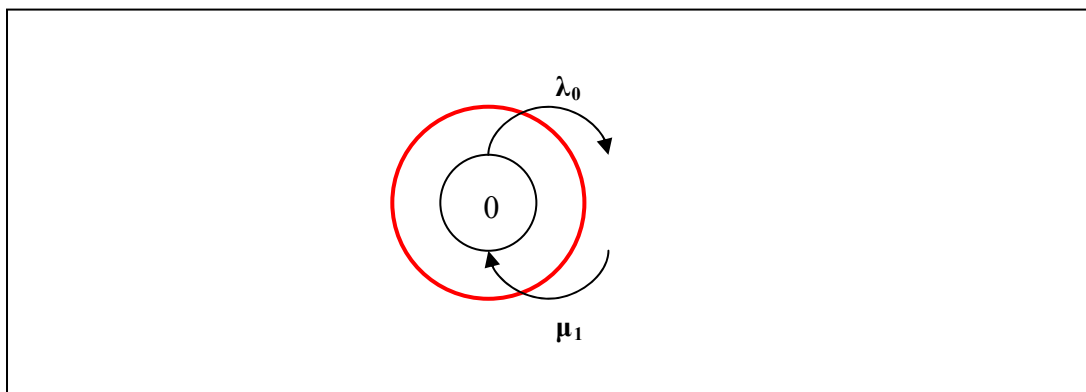


Figura 3.10. Bilanciamento dei flussi entranti ed uscenti allo stato zero

un'uguaglianza tra flusso entrante ($p_0 \lambda_0$) e flusso uscente ($p_1 \mu_1$), per cui la probabilità di stato p_1 è essere espressa come:

$$p_1 = \frac{\lambda_0}{\mu_1} p_0 \quad (1.47)$$

Analogamente, se $n \geq 1$, graficamente si ottiene il seguente bilanciamento (**Fig. 11**) del flusso entrante in un generico stato i ($p_{i+1} \mu_{i+1} + p_{i-1} \lambda_{i-1}$) ed uscente ($p_i (\lambda_i + \mu_i)$) da cui

$$p_2 = \frac{\lambda_1}{\mu_2} p_1 = \frac{\lambda_1 \lambda_0}{\mu_2 \mu_1} p_0$$

$$p_3 = \frac{\lambda_2}{\mu_3} p_2 = \frac{\lambda_2 \lambda_1 \lambda_0}{\mu_3 \mu_2 \mu_1} p_0$$

...

$$p_n = \frac{\prod_{i=0}^{n-1} \lambda_i}{\prod_{i=1}^n \mu_i} p_0 \quad (1.48)$$

che è l'espressione generale che fornisce p_n in funzione di p_0 .

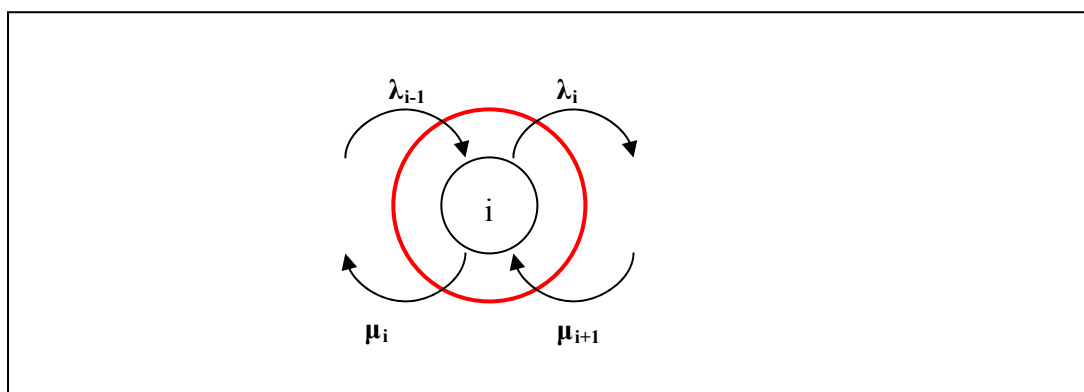


Figura 3.11. Bilanciamento dei flussi entranti ed uscenti in un generico stato i

Considerando l'espressione (1.45) risulta che $\sum_{n=1}^{\infty} p_n + p_0 = 1$, e, quindi, si può ricavare l'espressione di

p_0 :

$$p_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \frac{\prod_{i=0}^{n-1} \lambda_i}{\prod_{i=1}^n \mu_i}} \quad (1.49)$$

Condizione necessaria affinché ci sia esistenza dello stato stazionario è che

$$\sum_{n=1}^{\infty} \frac{\prod_{i=0}^{n-1} \lambda_i}{\prod_{i=1}^n \mu_i} < \infty \quad (1.50)$$

poiché se fosse infinita si avrebbe $p_n = 0$ per ogni n finito, e quindi lo stato del processo crescerebbe infinitamente e non si raggiungerebbe l'equilibrio e, quindi, la distribuzione stazionaria. [21]

Applicazioni dei processi di nascita e morte verranno riportate nel capitolo 2.

Nel caso $\mu_i = 0$ e $\lambda_i = \lambda \quad \forall i$ si ha un **processo di pura nascita**.

Assumendo che il sistema inizia al tempo t_0 con zero membri, ovvero

$$p_i(0) = \begin{cases} 1 & i=0 \\ 0 & i \neq 0 \end{cases} \quad (1.51)$$

le relazioni di Kolmogorov (1.43) diventeranno in questo caso

$$\begin{cases} dp_0(t) / dt = - p_0(t) \lambda \\ dp_i(t) / dt = - p_i(t) \lambda + p_{i-1}(t) \lambda \end{cases} \quad (1.52)$$

Nel caso inverso in cui $\mu_i = \mu$ e $\lambda_i = 0 \quad \forall i$ si avrà un *processo di pura morte* e si giungerà ad una distribuzione di Erlang (vedi par. 1.5.1):

$$dp_0(t) / dt = \mu e^{-\mu t} \frac{(\mu t)^{n-1}}{(n-1)!} \quad n = 0. \quad [19] \quad (1.53)$$

3.6 Misure Di Prestazione Dei Principali Sistemi Di Code

Nel capitolo precedente sono state indicate le grandezze principali, da tenere in considerazione per l'analisi di un sistema di servizio in condizioni stazionarie e sono state riassunte nella **tabella 3**.

Nel caso in cui i modelli sono Markoviani, si utilizzano, per modellare e calcolare tali misure di prestazione, i processi di nascita e morte e i concetti ritrovati a tal proposito nel paragrafo 1.6.

Conoscendo il valore della probabilità che nessun cliente sia nel sistema (\mathbf{p}_0) dall'espressione (1.49), riscrivibile come:

$$p_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \prod_n} \quad (2.1)$$

si può ricavare la relazione (1.48) della probabilità che lo stato del sistema sia n (\mathbf{p}_n) in funzione di \mathbf{p}_0 che può essere riscritta in questo modo:

$$p_n = \prod_n p_0 \quad (2.2)$$

dove

$$\prod_n = \frac{\prod_{i=0}^{n-1} \lambda_i}{\prod_{i=1}^n \mu_i} \quad (2.3)$$

La condizione necessaria affinché il sistema fosse a regime è rappresentata dalla convergenza di tale sommatoria

$$\sum_{n=1}^{\infty} \prod_n < \infty \quad (2.4)$$

Una volta calcolato il valore p_n attraverso la relazione precedente, possono essere calcolate le misure di prestazione N ed N^q attraverso le relazioni (1.6) e (1.7). [19]

Conoscendo tali grandezze, si possono, poi, individuare i valori di T e T^q applicando la legge di Little e, quindi le relazioni (1.17) e (1.18), prestando attenzione al valore di λ che deve rappresentare la frequenza media degli arrivi al sistema effettiva. Per cui la frequenza media degli arrivi coinciderà con il coefficiente di natalità se quest'ultimo è costante ($\lambda = \lambda_n$) mentre, nel caso in cui essa varia con lo stato n , è necessario calcolare la **frequenza media effettiva** $\hat{\lambda}$ attraverso la seguente relazione:

$$\hat{\lambda} = \sum_{n=0}^{\infty} \lambda_n p_n. \quad (2.5)$$

Quindi per effettuare un'**analisi stazionaria** dei sistemi di code si segue il seguente **percorso di calcolo** delle misure di prestazione:

1. **verifica della condizione necessaria** affinché ci sia esistenza dello stato stazionario espressa dalla relazione (2.4);
2. **calcolo** di p_0 utile per poter giungere al valore di p_n data dall'espressione (2.1);

3. *calcolo* di N ed N^q applicando la definizione di valore atteso di una v.a. discreta e quindi le relazioni (1.6) e (1.7);

4. *calcolo* della *frequenza media effettiva* $\hat{\lambda}$

$$\hat{\lambda} = \begin{cases} \lambda & \text{se } \lambda = \lambda_n & \text{per } n = 0, 1, \dots \\ \sum_{n=0}^{\infty} \lambda_n p_n & \text{se } \lambda_n = \lambda_{n(n)} & \text{per } n = 0, 1, \dots \end{cases} \quad (2.6)$$

5. *calcolo* delle misure di prestazione T e T^q attraverso la legge di Little

$$N = \hat{\lambda} T \quad (2.7) \quad N^q = \hat{\lambda} T^q. \quad (2.8)$$

3.6.1 Sistemi a singolo servente M/M/1

Sfruttando la notazione di Kendall, si può affermare che un sistema M/M/1 è caratterizzato da:

- una distribuzione degli intertempi di arrivo e dei tempi di servizio indipendenti e identicamente distribuiti secondo una distribuzione esponenziale, rispettivamente, di parametro λ e μ ;
- costituito da un unico server e, quindi, da un'unica coda (**Fig. 12**);
- capacità del sistema infinita;
- dimensione della popolazione infinita;
- disciplina di servizio di tipo FIFO.

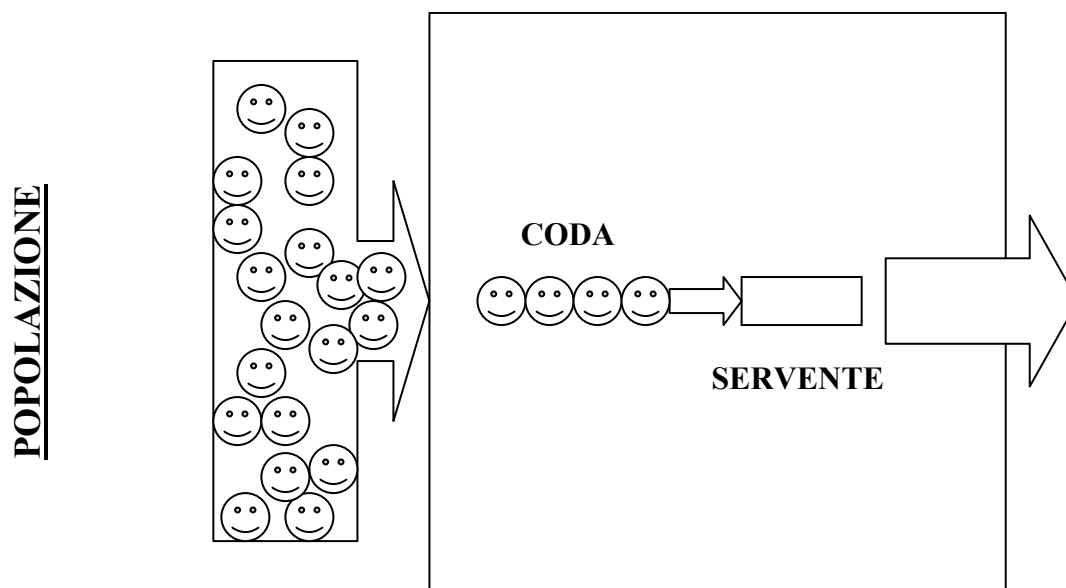


Figura 3.12. Esempio di sistema di servizio M/M/1

Seguendo il processo di calcolo delle prestazioni presentato nel precedente paragrafo, si individua la condizione necessaria di esistenza dello stato stazionario. A questo proposito si assume che il fattore di utilizzazione sia:

$$\rho = \frac{\lambda}{\mu} < 1 \quad (2.9)$$

detta **condizione di stabilità**. Si noti che un modello ad unico servente è caratterizzato da coefficienti di natalità e di mortalità costanti

$$\lambda_n = \lambda \quad \text{e} \quad \mu_n = \mu \quad n = 0, 1, 2, \dots$$

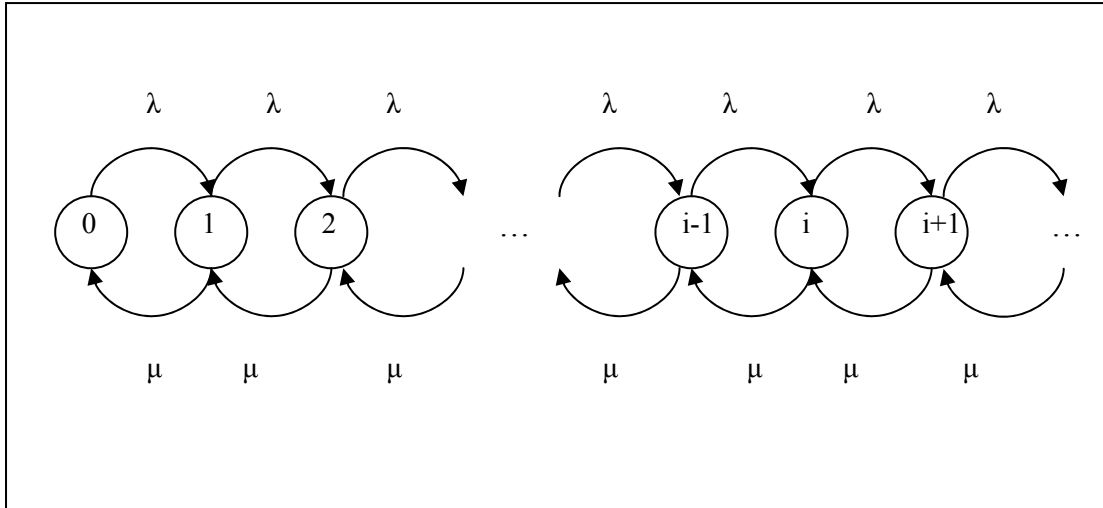


Figura 3.13. Processo di nascita e morte che rappresenta un sistema M/M/1

In questo modo risulta

$$\prod_n = (\lambda/\mu)^n = \rho^n \quad (2.10)$$

ed essendo $0 < \rho < 1$ si verifica l'effettiva convergenza della seguente serie

$$\sum_{n=1}^{\infty} \rho^n = \sum_{n=0}^{\infty} \rho^n - 1 = \frac{1}{1-\rho} - 1. \quad (2.11)$$

Si può, ora, calcolare la probabilità \mathbf{p}_0 utile per individuare il valore p_n :

$$\mathbf{p}_0 = \frac{1}{\frac{1}{1-\rho}} = 1 - \rho. \quad (2.12)$$

Dall'espressione (2.12) si può osservare che il tasso di utilizzazione risulta pari a:

$$\boldsymbol{\rho} = 1 - p_0 \quad (2.13)$$

e, quindi, può essere interpretato come il tasso di occupazione del servente, ossia la frazione di tempo in cui è impegnato (**Periodo di Occupazione**) ad espletare un servizio, che può essere visto come complemento del **Periodo di Inattività** (Fig. 14), che è l'intervallo di tempo che va dal momento in cui il servitore diventa inattivo all'istante in cui il servitore riceve la successiva richiesta. Essa può essere interpretata, anche, come la probabilità che ci sia almeno un cliente nel sistema, poiché il servente è unico e, quindi, ciò implicherebbe che sta offrendogli un servizio. [21]

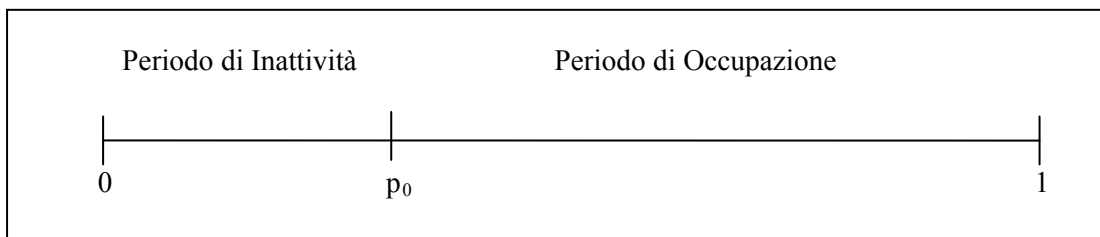


Figura 3.14. Periodo di Occupazione come complemento del Periodo di Inattività

Dal valore di p_0 si ricava la probabilità che lo stato del sistema sia n :

$$p_n = \rho^n (1 - \rho) \quad \text{per } n = 0, 1, 2, \dots \quad (2.14)$$

utilizzabile per calcolare le grandezze N ed N^q nel caso di un sistema Si avrà che in un sistema M/M/1:

$$N = \frac{\lambda}{\mu - \lambda} \quad [19] \quad (2.15)$$

N^q risulterà pari a:

$$N^q = \frac{\lambda^2}{\mu(\mu - \lambda)} \quad (2.16)$$

Applicando, poi, il Teorema di Little e le relazioni (1.16) e (1.17) si otterrà il valore delle altre grandezze:

$$T = \frac{1}{\mu - \lambda} \quad (2.17)$$

e

$$\mathbf{T}^q = \frac{\lambda}{\mu(\mu - \lambda)}. \quad [18] \quad (2.18)$$

Nella **tabella 4** si effettua un riepilogo delle misure di prestazione per un sistema M/M/1.

Misure di prestazione per un sistema M/M/1	
Probabilità che lo stato del sistema sia n	$\mathbf{p}_n = \rho^n (1 - \rho)$
Valore atteso dello stato del sistema (n)	$\mathbf{N} = \frac{\lambda}{\mu - \lambda}$
Valore atteso del numero di clienti in coda (n ^q)	$\mathbf{N}^q = \frac{\lambda^2}{\mu(\mu - \lambda)}$
Valore atteso del tempo passato nel sistema (t ^a)	$\mathbf{T} = \frac{1}{\mu - \lambda}$
Valore atteso del tempo passato in coda (t ^q).	$\mathbf{T}^q = \frac{\lambda}{\mu(\mu - \lambda)}$

Tabella 3.2. Riepilogo delle misure di prestazione per un sistema M/M/1

3.6.2 Sistemi multiserventi M/M/s

Sfruttando la notazione di Kendall, si può affermare che un sistema M/M/s è caratterizzato da:

- una distribuzione degli intertempi di arrivo e dei tempi di servizio indipendenti e identicamente distribuiti secondo una distribuzione esponenziale;
- costituito da s serventi che assumiamo lavorano in parallelo (**Fig. 3**);
- capacità del sistema infinita;
- dimensione della popolazione infinita;
- disciplina di servizio di tipo FIFO.

Questi modelli sono casi particolari dei processi di nascita e morte quando il sistema presenta un solo servente, $s = 1$. Infatti, come si è visto nel paragrafo precedente un sistema M/M/1 è rappresentabile mediante un processo di nascita e morte con coefficienti di natalità e di mortalità costanti.

Nel caso in cui $s > 1$ il sistema è rappresentabile con un processo di nascita e morte con coefficiente di natalità costante $\lambda_n = \lambda$, ma in cui il coefficiente di mortalità μ_n , che rappresenta la velocità media complessiva con cui vengono espletati i servizi e che determina il processo di partenza degli utenti dal sistema, risulta essere diverso a seconda di due casi. Considerando che la velocità media di servizio di ogni server è pari a μ , se gli utenti nel sistema sono $n \leq s$ il processo di partenza sarà caratterizzato da n variabili esponenziali μ pari al numero di utenti nel sistema; se, invece il numero di utenti supera il numero di serventi, se $n > s$, si avranno s variabili esponenziali poiché i server occupati continueranno ad essere sempre s :

$$\mu_n = \begin{cases} n\mu & \text{per } 1 \leq n \leq s \\ s\mu & \text{per } n \geq s \end{cases} \quad (2.19)$$

Il processo di nascita e morte che rappresenta il sistema a coda M/M/s è riportato in **figura 15**.

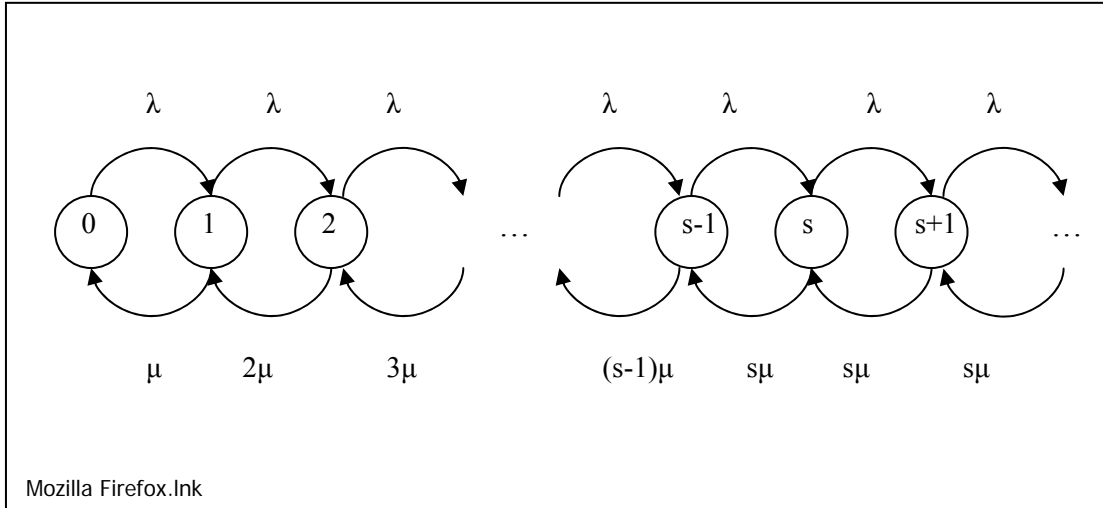


Figura 3.15. Processo di nascita e morte che rappresenta un sistema M/M/s

Per un sistema di questo tipo si può assumere che il fattore di utilizzazione sia $\rho = \lambda/(s\mu) < 1$. In questo modo ci si assicura l'esistenza dello stato stazionario. Infatti, si avrà

$$\Pi_n = \begin{cases} \frac{1}{n!} \left(\frac{\lambda}{\mu} \right)^n & \text{per } n = 1, \dots, s-1 \\ \frac{1}{s!} \left(\frac{\lambda}{\mu} \right)^s \left(\frac{\lambda}{s\mu} \right)^{n-s} & \text{per } n = s, s+1, \dots \end{cases} \quad (2.20)$$

La condizione di esistenza dello stato stazionario sarà

$$\sum_{n=1}^{\infty} \Pi_n = \sum_{n=1}^{s-1} \frac{1}{n!} (s\rho)^n + \frac{1}{s!} (s\rho)^s \frac{1}{1-\rho} < \infty \quad [19] \quad (2.21)$$

La probabilità che nessun cliente sia nel sistema (\mathbf{p}_0) e quella relativa alla probabilità che ce ne siano n (\mathbf{p}_n) saranno [21]:

$$\mathbf{p}_0 = \frac{1}{\sum_{n=0}^{s-1} \frac{1}{n!} (s\rho)^n + \frac{1}{s!} (s\rho)^s \frac{1}{1-\rho}} \quad (2.22)$$

$$\mathbf{p}_n = \begin{cases} \frac{1}{n!} \left(\frac{\lambda}{\mu} \right)^n p_0 & \text{per } n = 1, 2, \dots, s-1 \\ \frac{1}{s!} \left(\frac{\lambda}{\mu} \right)^s \left(\frac{\lambda}{s\mu} \right)^{n-s} p_0 & \text{per } n = s, s+1, \dots \end{cases} \quad (2.23)$$

Quest'ultima può essere utilizzata per il calcolo di \mathbf{N}^q [20]:

$$\mathbf{N}^q = \frac{1}{s!} \left(\frac{\lambda}{\mu} \right)^s \frac{\rho}{(1-\rho)^2} p_0 \quad (2.24)$$

Dalla Legge di Little si ricavano i valori delle grandezze \mathbf{N} , \mathbf{T} e \mathbf{T}^q :

$$\mathbf{T}^q = \mathbf{N}^q / \lambda \quad (2.25)$$

$$\mathbf{T} = \mathbf{T}^q + \frac{1}{\mu} \quad (2.26)$$

$$\mathbf{N} = \lambda \mathbf{T} \quad (2.27)$$

Nella **tabella 3** si riepilogano le misure di prestazione per un sistema M/M/s.

Misure di prestazione per un sistema M/M/s	
Probabilità che lo stato del sistema sia n	$p_n = \begin{cases} \frac{1}{n!} (s\rho)^n p_0 & \text{per } n = 1, \dots, s-1 \\ \frac{s^s}{s!} \rho^n p_0 & \text{per } n = s, s+1, \dots \end{cases}$
Valore atteso dello stato del sistema (n)	$N = \lambda T$
Valore atteso del numero di clienti in coda (n^q)	$N^q = \frac{1}{s!} \left(\frac{\lambda}{\mu} \right)^s \frac{\rho}{(1-\rho)^2} p_0$
Valore atteso del tempo passato nel sistema (t^q)	$T = T^q + \frac{1}{\mu}$
Valore atteso del tempo passato in coda (t^q).	$T^q = N^q/\lambda$

Tabella 3.3. Riepilogo delle misure di prestazione per un sistema M/M/s

3.7 Introduzione Al Modello Palm/Erlang-A.

Nel secondo capitolo sono stati proposti diversi modelli, rappresentanti ognuno sistemi di code con diverse caratteristiche, in cui si è ipotizzato che gli utenti in arrivo al sistema si unissero in coda, attendessero di essere serviti e lasciassero il sistema una volta che il servizio richiesto fosse terminato. Con tali assunzioni, implicitamente, si è considerata una pazienza del cliente infinita tale che, qualsiasi

fosse il tempo di attesa offerto dal sistema, egli non lo avrebbe abbandonato fin quando il suo servizio non fosse completato (**Fig.16**).

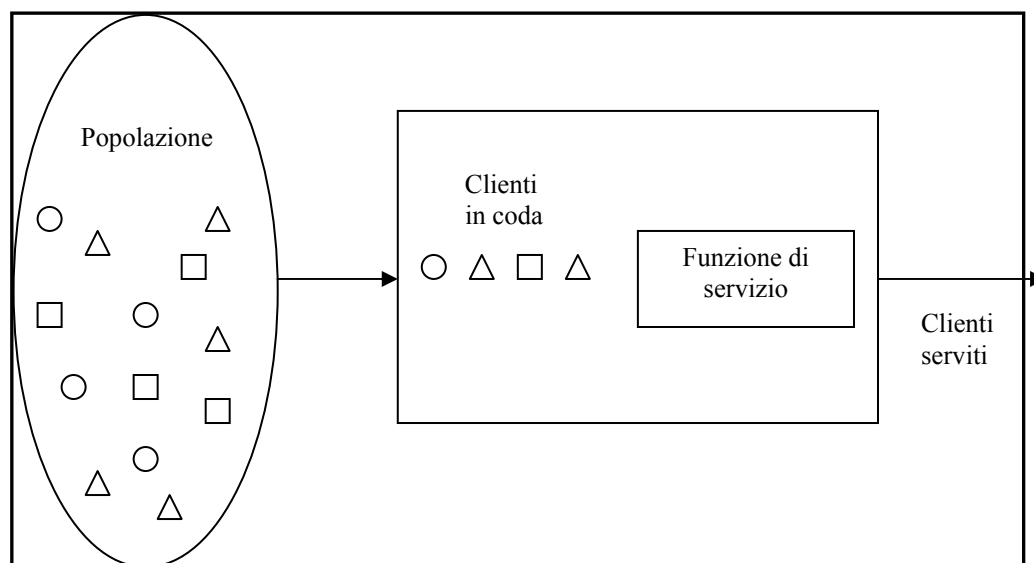


Figura 3. 16. Sistema di servizio senza alcuna forma di abbandono dei clienti

Non sempre un cliente attende la fine del servizio. Infatti, se ritiene che il tempo di attesa offerto non sia per lui appropriato, egli potrebbe abbandonare il sistema.

In **Fig. 17** si è rappresentato il *diagramma di flusso* (o *blue print*) attraverso il quale si può ben evidenziare il comportamento di un cliente in un sistema di servizio.

Si ipotizzi, come semplificazione, che ci sia una sola coda nel sistema e che il cliente non abbia la possibilità di jockeying, ossia di passare da una coda ad un'altra nella speranza di ridurre il proprio tempo di attesa. All'arrivo di un cliente al sistema, se non c'è alcun servente libero per poter usufruire del servizio, il cliente può ritenere opportuno abbandonare il sistema immediatamente o in seguito, valutando la lunghezza della coda o il tempo di attesa offertogli, cioè il tempo che dovrebbe attendere nel caso in cui la sua pazienza fosse infinita. Se il cliente ritiene che il tempo di attesa offerto sia troppo lungo, abbandona il sistema, altrimenti aspetta di essere servito.

Le modalità di abbandono che un utente può assumere sono:

-balking: il cliente decide di non unirsi alla coda perché non è disposto ad attendere il tempo offerto dal sistema;

-reneging: il cliente decide di unirsi alla coda ma la lascia in seguito poiché dopo un certo tempo non ha ancora ricevuto il servizio.

Il reneging può essere considerato l'abbandono per eccellenza poiché l'utente effettivamente entra nel sistema e poi decide di lasciarlo; mentre il balking influenza, principalmente il processo di arrivo dei clienti al sistema, ossia, solo una parte di essi si unirà alla coda (**Fig. 21**).

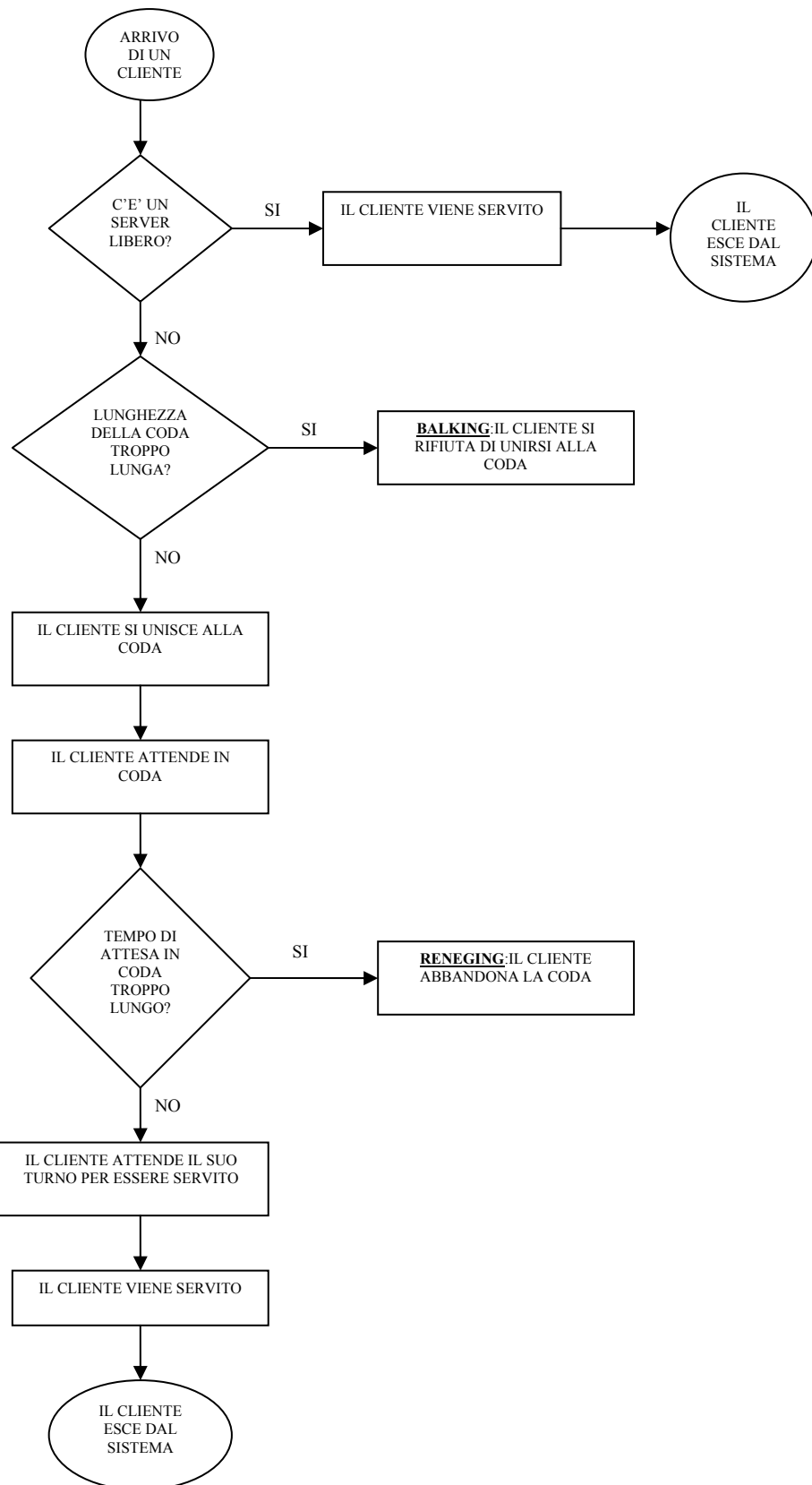


Figura 3.17. Blue print del comportamento del cliente in un sistema di servizio

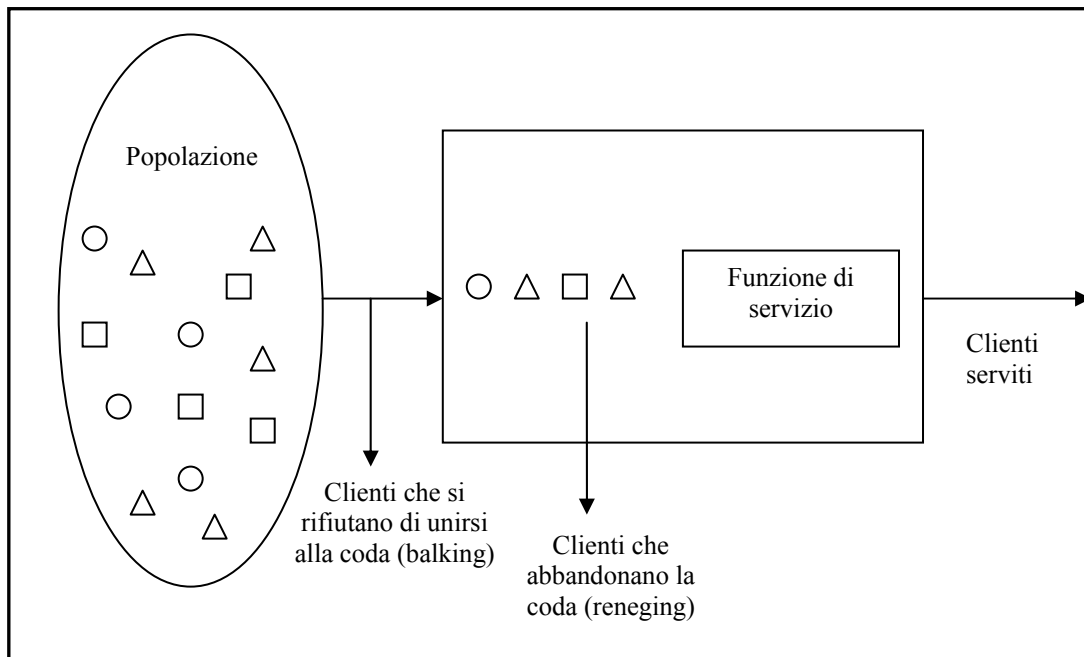


Figura 3.18. Sistema di servizio con balking e reneging

Un modello trattabile che ingloba possibilità di abbandono da parte del cliente fu introdotto da Palm nel 1957. Egli propose di estendere il modello M/M/s in M/M/s+M, associando ad ogni cliente in arrivo un tempo di pazienza esponenzialmente distribuito con media $1/\theta$ (“+M” indica tale caratterizzazione esponenziale dei tempi di abbandono). Formalmente, possiamo dire che in un sistema caratterizzato da tale modello, i clienti arrivano attraverso un processo di Poisson di parametro λ e con tempi di pazienza τ indipendenti ed identicamente distribuiti (i.i.d.) esponenzialmente con parametro θ . Vengono serviti attraverso una logica FIFO e con tempi di servizio i.i.d. in maniera esponenziale con media $1/\mu$.

L’ M/M/s+M, denominato **modello Palm/Erlang-A** o, semplicemente, **Erlang-A**, è, quindi, caratterizzato da:

- λ , tasso di arrivo
- μ , tasso di servizio ($1/\mu$ è la durata media del servizio)
- s , numero di serventi (considerati omogenei e in parallelo tra loro)

che sono parametri comuni al modello M/M/s. E, in più, bisogna considerare:

- θ , tasso di abbandono dei clienti dal sistema ($1/\theta$ è la pazienza media di un cliente).

Nella **Tab. 3.4** è possibile confrontare i modelli M/M/s e M/M/s+M per capire l'importanza dell'abbandono in un sistema di servizio. Si nota che una pazienza esponenziale dei clienti con media pari a 2 minuti dà origine al 3,1% di abbandoni e il tempo di attesa medio e la lunghezza media della coda sono minori rispetto al modello M/M/s. Questo è dovuto ad una riduzione del livello di congestione del sistema dato proprio dall'abbandono.

	M/M/s	M/M/s+M	M/M/s, $\lambda \downarrow 3,1\%$
Frazione di abbandoni	-	3,1%	-
Tempo medio di attesa	20,8 s	3,7 s	8,8 s
Lunghezza media della coda	17	3	7

Tabella 3.4. Confronto tra i modelli con e senza abbandono
50 agenti, $\lambda=48$, $1/\mu=1\text{min}$, $1/\theta=2\text{min}$

Nello studio di un sistema di code in cui è previsto la possibilità del cliente di lasciare il sistema anche senza aver ricevuto il servizio richiesto, si potrebbe pensare di considerare un modello M/M/s, di minor complessità rispetto all'M/M/s+M. In **Tab.4** si può osservare che considerando un modello senza abbandono con un tasso di arrivo ridotto del 3,1%, si avranno valori minori di attesa anche in questo caso, poiché ci sarà comunque un sovraccarico minore del sistema dovuto ad un numero minore di clienti in arrivo. Nonostante ciò, non si raggiungeranno i livelli di prestazione di un modello M/M/s+M. Da qui, notiamo l'importanza nel considerare il parametro θ nello studio di un sistema di code e, quindi, l'importanza del modello Erlang-A. [1]

3.7.1 La pazienza del cliente in un sistema di code

Per un dato cliente, nel modello Erlang-A, oltre alle variabili casuali intertempo di arrivo t^a e tempi di servizio t^s , comuni al modello M/M/s, va considerata la variabile aleatoria τ , che rappresenta il tempo di *pazienza di un cliente*. Tale pazienza può essere considerata come il tempo massimo che un cliente è disposto ad aspettare per il servizio oltre il quale, se non lo ha ancora ricevuto, abbandona la coda.

Nel modello M/M/s+M, τ è distribuita esponenzialmente con parametro θ :

$$\theta = \frac{1}{E[\tau]}. \quad (3.1)$$

Come per una qualsiasi v.a., una descrizione base dei tempi di pazienza di un cliente τ , è rappresentata dalla sua funzione di distribuzione cumulata F . Una equivalente descrizione è la sua corrispondente funzione di rischio che descrive la frequenza con la quale si determina un insuccesso del sistema, in questo caso, un abbandono di un cliente da esso. Assumendo che F abbia una densità $f = \frac{dF}{dt}$, la funzione di rischio sarà

$$h(t) = \frac{f(t)}{[1 - F(t)]} \quad \text{con } t \geq 0. \quad (3.2)$$

$[1 - F(t)]$ è la probabilità che non avvenga un abbandono, detta anche probabilità di sopravvivenza, in genere pari a

$$1 - F(t) = S(t) = \Pr \{T > t\} \quad (3.3)$$

dove T è un tempo limite di sopravvivenza (il tempo di pazienza del cliente) e t è un istante temporale di riferimento (il tempo di attesa offerto dal sistema di servizio).

Di conseguenza, la probabilità cumulata dei tempi di abbandono sarà

$$F(t) = 1 - S(t) = \Pr \{T \leq t\} \quad (3.4)$$

Intuitivamente, $h(t) dt$ è la probabilità che un cliente, che ha atteso già per t unità di tempo, abbandonerà il sistema nelle prossime dt unità, quindi durante l'intervallo $(t, t+dt)$, ossia

$$h(t) dt = \Pr \{t \leq T < t+dt \mid T > t\} \quad (3.5)$$

Quindi, la probabilità cumulata della funzione di rischio nell'intervallo suddetto sarà:

$$H(t) = \int_t^{t+dt} h(t) dt = \frac{F(t)}{S(t)} \quad (3.6)$$

ed è pari al rapporto tra la probabilità complessiva di abbandoni dal sistema e la probabilità cumulata che i clienti attendono di essere serviti, in un certo intervallo temporale.

Date queste premesse, si può affermare che $h(t)$ offre una naturale descrizione dinamica della pazienza, come essa evolve durante l'attesa del servizio e può essere, quindi, considerata come la funzione di (im) pazienza di un cliente.

Si è assunto che la pazienza del cliente sia esponenzialmente distribuita con parametro θ . Questo comporta che tale funzione di (im) pazienza $h(t)$ risulta essere costante:

$$h(t) = \frac{f(t)}{(1 - F(t))} = \theta. \quad (3.7)$$

Ciò non rispecchia l'andamento reale della funzione di pazienza di un cliente normale, meglio rappresentata da una sua distribuzione generica (**Fig.19**) .

Ma si è dimostrato che considerare un modello M/M/s+M è, comunque, una buona approssimazione del modello M/G/s+G, caratterizzato da tempi di servizio e di abbandono distribuiti con distribuzione generica, assunzioni molto più vicine alla realtà. [16] [22]

L'assunzione markoviana è preferibile poiché facilita, come si è visto negli scorsi capitoli, lo studio del sistema di code tramite processi di nascita e morte ed equazioni di bilancio per il calcolo delle misure di prestazione.

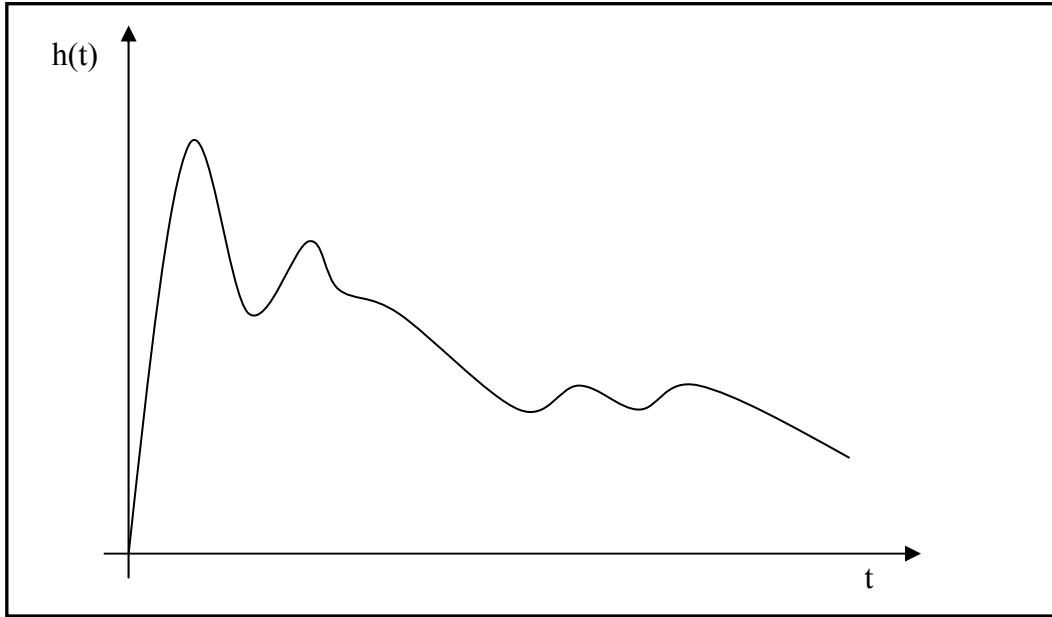


Figura 3.19. Esempio di distribuzione della funzione di pazienza di un cliente

Esiste, inoltre, una relazione lineare importante tra la frazione di abbandoni e il ritardo atteso utile per il calcolo del tasso di abbandono θ , valida nel caso in cui τ risulta essere una v.a. esponenziale:

$$P \{ \text{Abbandono} \} = \theta * E[\text{tempo di attesa}] = \theta * E[W]. \quad (3.8)$$

Infatti, considerando l'equazione di bilancio tra il tasso con cui i clienti che entrano nel sistema lo abbandonano e il tasso con cui i clienti abbandonano la coda

$$\lambda * P \{ \text{Ab} \} = \theta * E[\text{lunghezza della coda}] = \theta * [N^q] \quad (3.9)$$

e applicando la legge di Little (par. 1.4.4) si ottiene

$$\lambda * P \{ \text{Ab} \} = \theta * \lambda * E[W] \quad (3.10)$$

da cui, semplificando λ , si ricava la relazione (3.8).

W è l'attesa effettiva del cliente nel sistema pari al minimo tra tempo di attesa in coda offerto dal sistema V e il tempo di pazienza τ , ossia:

$$W = \min\{V, \tau\}. \quad (3.11)$$

Infatti:

- se $\tau > V$, il cliente attende di essere servito e, quindi, lascia il sistema quando il servizio è terminato, allora $W = V$;
- se $\tau < V$, il cliente abbandona il sistema prima di essere servito, allora $W = \tau$.

Grazie alla relazione precedente di cui gode il modello Erlang-A, il tempo medio di attesa del cliente $E[W]$ e la frazione dei clienti che la abbandonano $P\{Ab\}$, sono i dati di input attraverso i quali stimare il tasso di abbandono θ :

$$\hat{\theta} = \frac{1}{E[\tau]} = \frac{P\{Ab\}}{E[W]} = \frac{\%Abbandono}{attesa\ media} \quad (3.12)$$

parametro utile per caratterizzare tale modello.

Però, come indice empirico di pazienza, di solito, viene utilizzato il rapporto tra la percentuale di clienti serviti e di quella dei clienti che hanno abbandonato il sistema:

$$\text{Indice empirico di pazienza} = \frac{\%serviti}{\%abbandoni} \quad [1] \quad (3.13)$$

valori facilmente ricavabili dal database del sistema di servizio in esame.

3.8 Code Visibili E Code Invisibili: L'importanza Dell'informazione

L'attesa in un sistema di servizio provoca, in un cliente impaziente, differenti stati d'animo che lo portano ad assumere diverse decisioni e comportamenti riguardo alla sua permanenza nel sistema, a seconda che la coda sia a lui visibile o meno.

Si assuma che una *coda* sia *invisibile* quando gli utenti in attesa non ottengono alcuna informazione riguardante la lunghezza della coda o il loro tempo di attesa restante prima di ricevere il servizio. Esempi tipici di sistemi caratterizzati da una coda di questo tipo sono i call center o sistemi i cui servizi sono basati sull'utilizzo di Internet [17].

Quando il cliente non ha possibilità di avere alcuna informazione sui ritardi previsti o alcuno strumento che possa essergli utile per una tale valutazione, all'avanzare del suo tempo di attesa, la sua sensazione di disagio e la sua frustrazione cresce lentamente (Fig. 23).

Nel caso in cui informazione o strumenti tali da poter valutare il ritardo previsto o il proprio progresso in coda sono disponibili al cliente, la **coda** verrà considerata **visibile**. Esempi tipici sono uffici postali e banche.

In Fig. 20 si può notare che, in questo caso, lo stato d'animo del cliente diventa via via sempre più sereno. Essendo a conoscenza della sua posizione in coda e/o del suo tempo di attesa al trascorrere del tempo, può, infatti prendere le sue decisioni con maggior tranquillità.

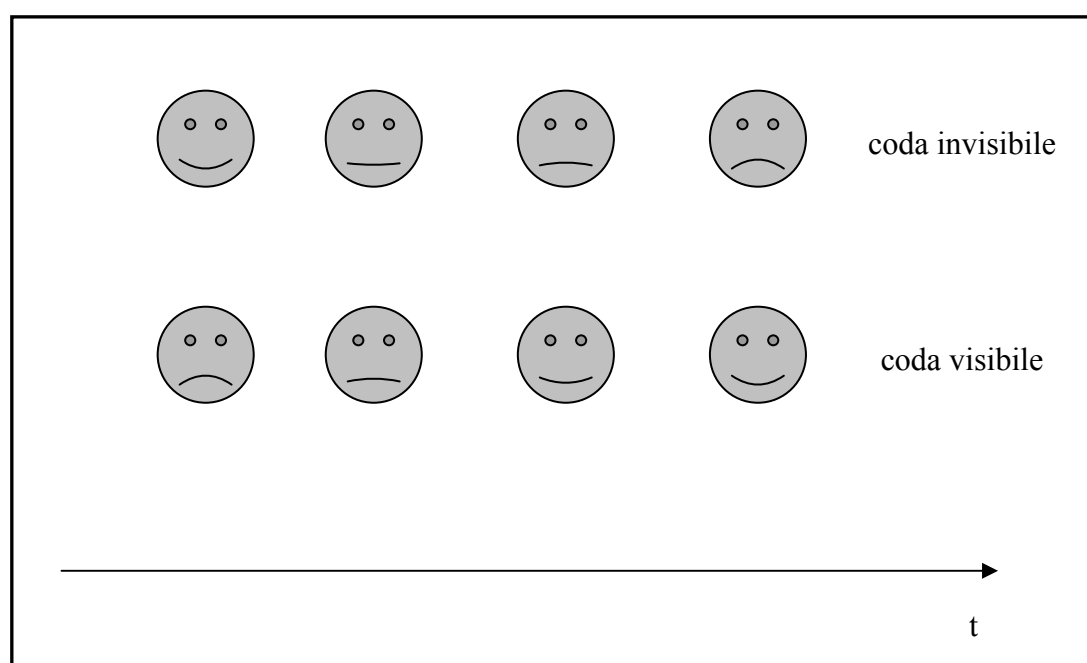


Figura 3.20. Stati d'animo di un cliente, nel tempo, in una coda visibile e una invisibile

L'esperienza di un cliente che si reca in un centro di contatto per richiedere un servizio è fondamentale per il gestore di tale centro, poiché da essa (e, ovviamente, anche da altri fattori) può dipendere il suo successo. Per cui, l'informazione sembra assumere una certa importanza, principalmente nel caso in cui la coda sia invisibile, sia per il gestore del servizio che per i clienti. Però, W.Feng nel 2003 nel suo lavoro *"Improving service for service systems with different arriving rate"*, simulando i due modelli e valutando il rendimento del sistema in entrambi i casi, evidenzia un rendimento maggiore del sistema nel caso di coda invisibile. Quindi per il manager risulta conveniente non dare informazioni al cliente, che rappresentano una sorta di incentivo all'abbandono dal sistema se le attese sono lunghe e, quindi un

guadagno minore. Mentre per i clienti è preferibile, ovviamente ricevere l'informazione anche perchè l'abbandono comporta tempi di attesa minori.

3.8.1 “Modello di Reneging” e “Modello di Balking”

Le due principali forme di abbandono di un cliente in un sistema di servizio, reneging e balking, si presentano con una probabilità differente in base alla tipologia della coda considerata.

Si è detto che se un server non è immediatamente disponibile, il cliente potrebbe lasciare immediatamente il sistema (balking). Si supponga che la probabilità che ciò avvenga sia pari a β . Viceversa, egli aspetta di essere servito con una probabilità pari a $1-\beta$. Tale fenomeno di balking è maggiormente presente nel caso in cui viene data al cliente, all'arrivo, informazione sulla lunghezza della coda o sui tempi di attesa, parametri utili per ponderare tale decisione di rifiuto.

Si consideri il caso in cui i clienti siano in una *coda invisibile* e non ricevono alcuna informazione sulla propria attesa. Non ricevendo, all'arrivo, informazioni o strumenti tali da poter decidere di abbandonare immediatamente il sistema, la probabilità di balking è ridotta al minimo e risulta maggiore quella di reneging. Per cui, un modello di questo tipo si può definire “*modello di reneging*”. Il tasso di arrivo sarà così caratterizzato:

$$\lambda_n = \begin{cases} \lambda & \text{per } 0 \leq n \leq s-1 \\ \lambda (1-\beta) & \text{per } 1 \leq n \leq s \end{cases} \quad (3.14)$$

dove $\lambda (1-\beta)$ è la porzione di cliente che arriva al sistema.

Quindi, se un server non è immediatamente disponibile e il cliente non lascia il sistema, potrebbe abbandonare la coda in seguito (reneging), dopo un tempo esponenziale con media $1/\theta$, se non ha ancora ricevuto il servizio. Altrimenti attende, riceve il servizio e esce dal sistema una volta che esso sia completato. Per cui il coefficiente di mortalità sarà:

$$\mu_n = \begin{cases} n\mu & \text{per } 1 \leq n \leq s \\ s\mu + (n-s)\theta & \text{per } n \geq s+1 \end{cases} \quad (3.15)$$

dove $s\mu + (n-s)\theta$ è il numero di clienti in uscita pari alla somma tra clienti serviti e clienti che hanno deciso di abbandonare la coda (reneging). [15]

Si è detto che dare informazioni al cliente al suo arrivo comporta una probabilità maggiore che egli ritenga immediatamente opportuno non aspettare per il servizio (balking), caso molto diffuso in sistemi a coda visibile.

Sia T il valore, all'equilibrio, del tempo di pazienza τ , oltre il quale il cliente abbandona il sistema.

Se T eccede il valore casuale del tempo di attesa offerto V_n (dipendente dallo stato del sistema n), ora conosciuto dal cliente, egli si unirà alla coda. Viceversa si rifiuterà di entrare nel sistema.

Assumendo che nessuno prima di lui abbandona la coda, se al suo arrivo un cliente trova $s+n$ clienti nel sistema (non contando se stesso) si unisce ad essa con probabilità

$$q_n = P(T > V_n) \quad n \geq 0. \quad (3.16)$$

V_n è una v.a. con distribuzione data dalla somma di $n+1$ esponenziali ognuno con media $1/s\mu$.

Di solito si considera nel calcolare q_n la media di V_n ottenendo la relazione:

$$\bar{q}_n = P(T > E[V_n]) = e^{-\alpha(n+1)/s\mu} \quad [15]. \quad (3.17)$$

Nel caso di **coda visibile** si avrà un modello che definiremo “**modello di balking**” in cui il tasso di arrivo sarà:

$$\lambda_n = \begin{cases} \lambda & \text{per } 0 \leq n \leq s-1 \\ \lambda (1-\beta) q_{n-s} & \text{per } n \geq s+1. \end{cases} \quad (3.18)$$

dove $\lambda (1-\beta) q_{n-s}$ è la porzione di clienti che all'arrivo decide di unirsi alla coda.

Sarà, invece, caratterizzato da un tasso di mortalità:

$$\mu_n = \begin{cases} n\mu & \text{per } 1 \leq n \leq s \\ s\mu & \text{per } n \geq s+1. \end{cases} \quad (3.19)$$

dove i clienti che escono dal sistema sono semplicemente quelli che vengono serviti, essendo, in questo caso, la possibilità di reneking ridotta al minimo. Formalmente si può dire che nel modello di balking il cliente, ricevendo all'arrivo informazioni sulla sua attesa, decide subito cosa fare e, una volta unitosi alla coda, non la abbandona fino a che non viene servito. [15]

Entrambi i modelli discendono da un **modello generale** che ingloba entrambe le possibilità: quella di balking e quella di reneking.

Si consideri che un cliente, con $j-1$ utenti prima di lui in coda, la abbandona con un tasso δ'_j . Il tasso totale di reneking quando ci sono $s+n$ clienti nel sistema sarà:

$$\delta_n = \sum_{j=1}^n \delta'_j \approx \sum_{j=1}^n h(j/\lambda) \quad \text{per } n \geq 0. \quad (3.20)$$

dove $h(x)$ è la funzione di rischio definita nel paragrafo 3.1.1.

I tassi di arrivo e quello di servizio saranno:

$$\lambda_n = \begin{cases} \lambda & \text{per } 0 \leq n \leq s-1 \\ \lambda (1-\beta) q_{n-s} & \text{per } n \geq s+1. \end{cases} \quad (3.21)$$

e

$$\mu_n = \begin{cases} n\mu & \text{per } 1 \leq n \leq s \\ s\mu + \delta_{n-s} & \text{per } n \geq s+1 \end{cases} \quad (3.22)$$

Ponendo $q_{n-s}=1$ e $\delta_{n-s} = (n-s) \theta$, si ottiene il modello di reneking in cui la coda risulta essere invisibile. Con $\delta_{n-s} = 0$, si ricava il modello di balking con coda visibile. [15]

CAPITOLO IV

PROTECTION LEVEL E TECNICHE DI OVERBOOKING. CASO APPLICATIVO.

Lo yield management è una tecnica gestionale che cerca di fornire, in modo rapido ma preciso, informazioni preziose ad aziende di servizi, caratterizzate da una domanda non determinabile con precisione e da risorse da gestire in numero limitato.

L'obiettivo è quindi quello di gestire in modo più profittevole possibile le risorse di cui si dispone. A tal proposito esistono due fasi di analisi: una prima che cerca di trovare il giusto equilibrio di posti da offrire a tariffe ridotte e posti da vendere a tariffa intera ed una seconda, nota come overbooking, orientata alla definizione del limite massimo di prenotazioni in eccesso rispetto alla capacità di cui si dispone per far fronte ad eventuali mancate fruizioni del servizio prenotato precedentemente.

L'azienda alla quale si è deciso di applicare lo yield management è un'associazione di ricerca e sviluppo, la ARES, che presenta tutti i requisiti per l'implementazione della tecnica in quanto offre corsi di formazione con posti in numero limitato, la cui domanda non è prevedibile in alcun modo anticipatamente.

All'associazione è stato applicato lo yield management focalizzandosi alla definizione del limite di overbooking, e per determinare tale valore sono state implementate due tecniche alternative e ne sono stati confrontati i risultati.

4.1 Introduzione

Le imprese di servizi, in modo particolare negli ultimi anni, si trovano a dover gestire una forte concorrenza in un mercato, caratterizzato da una elevata imprevedibilità dell'andamento della domanda, una rigidità della struttura produttiva in termini di capacità di erogazione del prodotto/servizio ed una prevalenza dei costi fissi.

Questo scenario ha indotto lo studio di strumenti capaci di supportare nel miglior modo possibile lo svolgimento del processo di gestione ed il raggiungimento degli obiettivi sia di breve che di medio-lungo periodo.

Il tentativo di risolvere tali problematiche ha determinato il sorgere di numerose tecniche gestionali innovative, tra le quali si evidenzia lo yield management.

Con tale espressione ci si riferisce ad una serie di metodologie, più o meno sofisticate, che rendono possibile la massimizzazione del rendimento (*yield*) attraverso l'utilizzo di modelli matematico-statistici che tengono conto contemporaneamente delle particolarità relative alla capacità produttiva dell'azienda (offerta) e di quelle relative al comportamento differenziato della clientela (domanda) [3]. Lo yield management opera quindi realizzando un arbitraggio dinamico tra domanda ed offerta [4].

In sintesi, l'ottimizzazione dei ricavi può essere ottenuta tramite una sistematica attività di ricerca e di cattura dell'intera domanda potenziale che l'azienda è in grado di soddisfare, minimizzando il rischio di mancati ricavi, individuando nuovi segmenti di domanda da servire ed infine sfruttando la capacità di contribuzione di ciascun segmento di clientela [2].

Il concetto di yield, precedentemente introdotto, può essere rappresentato dal seguente rapporto:

$$Yield = \frac{Rr}{Rp}$$

dove Rr individua il ricavo realizzato effettivamente dall'impresa, mentre Rp indica il ricavo potenziale ossia il ricavo ottenibile dall'impiego completo della capacità produttiva [1]. E' utile precisare che entrambi i ricavi si riferiscono al medesimo intervallo temporale. Il rapporto di yield è chiaramente compreso tra zero ed uno ed è proprio pari ad uno nel caso di saturazione nell'utilizzo della capacità produttiva dell'impresa.

In pratica le metodologie utilizzate ponderano e bilanciano convenientemente le peculiarità insite nella struttura produttiva con le esigenze espresse dai consumatori. Questa tecnica gestionale presuppone pertanto l'esistenza di una domanda governata da leggi probabilistiche gestibile attraverso una manovra sui prezzi associata ad adeguate tecniche di marketing [5].

La tecnica, inoltre non è statica ma si evolve in concomitanza dei cambiamenti del mercato ed infine non è unica e standardizzabile in quanto la si può applicare a diversi livelli di dettaglio a seconda delle esigenze dell'impresa [6].

4.2 Requisiti necessari per l'applicazione dello yield management

L'implementazione dello yield management non è semplice e richiede uno "scenario di riferimento" molto preciso, che lo rende applicabile solo in specifici contesti aziendali. In particolare, la tecnica è appropriata e consigliabile solo quando un'azienda possiede i seguenti requisiti [3] :

1. Offerta fissa

L'obiettivo generale dello yield management è l'allocazione efficiente della capacità in strutture nelle quali quest'ultima è fissa o comunque non modificabile nel breve periodo se non a costi talmente elevati da comprometterne definitivamente la convenienza economica. Proprio per tale ragione si parla di strutture con elevata rigidità produttiva.

2. Domanda segmentabile

I clienti hanno una diversa elasticità al prezzo¹ e pertanto vengono divisi in diversi segmenti di mercato. Per ogni segmento viene poi studiata una strategia di prezzo "su misura". Può pertanto accadere che il

¹ L'elasticità al prezzo è definita come la variazione percentuale della quantità domandata di un bene in seguito ad una variazione percentuale del prezzo.

medesimo servizio sia venduto a prezzi differenti a segmenti di clienti differenti in tempi e con metodi di prenotazione differenti.

3. Offerta non immagazzinabile

La mancata vendita di un prodotto/servizio genera una perdita che non potrà, in nessun modo, essere recuperata. Nello specifico se, nella giornata, i servizi potenzialmente erogabili non vengono venduti questi vanno definitivamente persi, per cui non si può sperare di venderli in futuro, immagazzinandoli, a differenza delle imprese che producono beni tangibili che, invece, possono creare scorte. Per tale ragione si parla di “beni deperibili”.

4. Prodotti/servizi vendibili in anticipo

La possibilità di vendere anticipatamente, tramite la prenotazione, rappresenta un fattore di sicurezza per le imprese. Infatti il vendere anche solo una parte della capacità disponibile in largo anticipo consente una gestione più efficiente. Questo sistema di vendita è tuttavia anche fonte di complicazioni: i fenomeni di *no-show* e delle cancellazioni hanno condotto il management verso politiche di overbooking, ossia politiche basate sull'accettazione di un numero di prenotazioni superiore a quelle che effettivamente il servizio è in grado di soddisfare.

5. Domanda variabile e stocastica

L'andamento della richiesta di un prodotto/servizio varia dinamicamente in modo incerto e molto spesso non prevedibile con grande precisione. L'andamento delle richieste è pertanto regolato da leggi probabilistiche.

6. Domanda influenzabile dalla strategia di prezzo

I consumatori devono poter essere influenzati dalla manovre di prezzo che, di volta in volta, vengono attuate dal management per fronteggiare l'andamento delle richieste di servizio.

7. Costi fissi nel breve termine

La struttura dei costi delle imprese di servizi studiate nell'ambito dello yield management, come già detto, è caratterizzata dalla preponderanza dei costi fissi che non possono essere diminuiti con alcun intervento nel breve periodo.

8. Bassi costi marginali di vendita

Una volta verificata la disponibilità di un'unità supplementare, la sua vendita viene realizzata a prezzi molto bassi, quasi esclusivamente di natura variabile.

9. Alti costi marginali di produzione

Sia che sia saturata la capacità sia che essa sia sfruttata in modo parziale ed incompleto, in ogni caso, l'azienda deve sostenere dei costi fissi di produzione molto elevati, dovuti alla sola erogazione del servizio

4.3 Applicazioni

Dopo aver elencato tutte le caratteristiche che deve possedere l'impresa che vuole applicare lo yield management è chiaro comprendere che questa tecnica gestionale non si adatta a tutte le imprese; in particolare quelle del settore terziario si adattano meglio ad essa.

I contesti applicativi in cui è documentabile una gestione attraverso la politica di yield management sono:

- Compagnie di trasporto aereo;
- Compagnie di trasporto ferroviario;
- Compagnie di trasporto marittimo;
- Aziende di noleggio di mezzi di trasporto;
- Aziende alberghiere;
- Aziende di noleggio di spazi espositivi in aree fieristiche;
- Tour operators;
- Navi da crociera;
- Ristoranti;
- Sale cinematografiche.

E' bene puntualizzare che questa è un'elencazione non esaustiva dei contesti applicativi della tecnica in quanto le implementazioni stanno aumentando velocemente negli ultimi anni e non sempre sono documentate.

In base al contesto applicativo dello yield management si può parlare di:

- **Seat allocation**

Consiste nel vendere lo stesso prodotto/servizio a tariffe diverse a seconda del momento in cui esso è acquistato. Questo tipo di gestione è tipica nel caso aereo, alberghiero o dei noleggi.

- **Intermediate leg management**

Consiste nel gestire in modo ottimo la disponibilità di posti tra due fermate successive ed è caratteristica del settore ferroviario. Nello specifico si tratta di rifiutare una prenotazione quando c'è una ragionevole certezza che quel posto possa essere venduto successivamente per una tratta più lunga, producendo in tal modo un maggior guadagno.

4.4 Strumenti dello yield management

Nell'ambito del processo di yield management si possono riconoscere quattro fasi principali: la segmentazione del mercato e il pricing², la previsione della domanda, l'allocazione della capacità produttiva, la prenotazione e la vendita [3].

4.4.1 Segmentazione del mercato e pricing

Il primo gradino da affrontare in un programma di yield management è la definizione dei vari segmenti di mercato ai quali rivolgersi. Successivamente, una volta individuati i vari segmenti, bisogna decidere come applicare loro differenti prezzi; tale politica è definita dagli economisti "discriminazione dei prezzi".

L'obiettivo è quello di espandere il più possibile il proprio potenziale *revenue* applicando le tariffe più elevate ai quei segmenti di mercato che non risentono dei cambiamenti dei livelli dei prezzi, cioè alle fasce di clienti che hanno una bassa elasticità della domanda al prezzo; di conseguenza si applicano poi le tariffe a minor margine contributivo a quei segmenti di mercato sensibili alla riduzione dei prezzi generando, in tal modo, un incremento delle vendite.

La massimizzazione dei ricavi si consegue, quindi, attraverso un'opportuna gestione del binomio prezzo-differenziazione.

Il modo più efficace per ottenere una buona segmentazione di mercato è identificare distinti gruppi di clienti che ragionano e si comportano in maniera diversa l'uno dall'altro e che sono di particolare interesse per quanto concerne le attività di marketing. Conoscere i comportamenti d'acquisto di ogni potenziale cliente risulta infatti fondamentale nella configurazione di un buon sistema di yield management.

Le variabili rispetto alle quali è effettuata la segmentazione della clientela sono numerose. Di certo, la principale variabile di segmentazione è la **motivazione**; bisogna però considerare che il medesimo cliente potrebbe avere nel tempo motivazioni differenti, pertanto, accanto a questa variabile si preferisce considerare una seconda variabile di segmentazione che è rappresentata dai **bisogni** che i clienti intendono soddisfare e dai benefici ricercati nella fruizione dello specifico prodotto/servizio. Già tramite la conoscenza delle suddette due variabili sarebbe possibile effettuare un buon posizionamento del prodotto/servizio sul mercato. Per ottenere un'ottima implementazione dello

² Con il termine *pricing* si intende la definizione della tariffa di un prodotto/servizio.

yield management è necessario però spingersi oltre e considerare una terza variabile di segmentazione: le **condizioni d'uso**.

In questo modo è consentito contenere i rischi connessi all'applicazione rigida degli assunti dello yield management, sulla base dei quali si può arrivare a rifiutare l'erogazione di un servizio ad un cliente o ad escluderlo, qualora questo non rientri nella classe aperta disponibile in quel momento, senza tenere conto del fatto che, magari, può trattarsi di un cliente abituale che garantisce ricavi distribuiti durante l'arco dell'anno o della stagione.

L'utilizzo di questa ulteriore discriminante consente, sulla base della frequenza d'uso, del tempo d'acquisto e della quantità acquistata, di far rientrare il cliente all'interno di una classe in considerazione della continuità del rapporto esistente.

Una volta effettuata una distinzione tra i vari clienti ed averli raggruppati in segmenti omogenei è necessario decidere per ogni segmento quale tariffa applicare. Si passa quindi alla seconda fase della discriminazione dei prezzi ossia il "*pricing*".

Il pricing è un'attività di marketing cui spesso viene data meno attenzione di quella che effettivamente meriterebbe. Nagle, esperto dell'argomento, commenta questo fatto dicendo che, sebbene il pricing venga sempre declamato come una delle 4P (price, place, product, promotion) del marketing mix³, nella pratica comune viene spesso trascurato in quanto, probabilmente, i marketing managers reputano il prezzo meno interessante e certamente meno centrale per l'ottenimento del successo rispetto alle strategie di prodotto, distribuzione o promozione. Fortunatamente, negli ultimi anni, i managers stanno iniziando a conferire maggiore attenzione alle politiche di pricing nel processo di marketing. In questo senso, è importante notare il ruolo delle nuove ricerche [4] nello sviluppo di strategie di prezzo che, in un secondo momento, saranno implementate in un sistema di yield management.

Ogni azienda affronta il pricing nel modo ritenuto più idoneo alla soluzione dei propri bisogni immediati, anche se, sempre, con un occhio al futuro.

Il pricing non è una funzione statica, bensì dinamica ed energica e le strategie di pricing devono essere utilizzate dalle imprese considerando i cambiamenti del tempo e le differenze nei prodotti. Le politiche di prezzo, pertanto, vengono costantemente riviste, aggiustate e manipolate in modo da risultare sempre allineate agli obiettivi dell'organizzazione; quindi se gli obiettivi cambiano, conseguentemente cambiano anche le politiche di pricing; non a caso si parla di *dynamic pricing*.

Le funzioni di pricing devono essere viste secondo una prospettiva interna e secondo una prospettiva esterna. I fattori interni che riguardano le strategie di marketing, gli obiettivi di marketing, i costi di

³ Con l'espressione *Marketing mix* si intende l'insieme di variabili di marketing, quali prodotto, promozione, prezzo e distribuzione, controllabili dall'impresa in modo diretto.

produzione e di vendita, interagiscono con i fattori esterni che, invece, si riferiscono alla natura del mercato, ai fattori che influenzano la domanda e l'offerta, alla competizione ed ad altri fattori ambientali. È difficile comprendere e determinare se il ruolo principale sia giocato dai fattori interni o dai fattori esterni, la cosa più probabile è che entrambi, in momenti differenti, giochino un ruolo importante nelle operazioni di business [7].

E' comunque certo che questo mix di fattori fornisce le basi per lo sviluppo della politica di pricing, che, senza dubbio, risulta essere la funzione di marketing più importante all'interno di un'organizzazione aziendale, infatti, senza politiche di prezzo efficienti, le aziende orientate al profitto non sopravvivrebbero. Il pricing viene spesso suddiviso in componenti strategiche al fine di favorirne l'implementazione. Tale processo richiede una profonda comprensione della sensibilità al prezzo del mercato o dei vari segmenti, nonché dei competitors e dei possibili effetti che le tattiche di prezzo potrebbero verosimilmente generare. Una volta stabiliti e studiati tutti i suddetti fattori è possibile iniziare lo sviluppo degli obiettivi strategici e delle tattiche che conducono alla formazione del prezzo o allo sviluppo della politica di prezzo.

Gli obiettivi di pricing possono essere suddivisi in tre diversi orientamenti:

- orientamento al profitto;
- orientamento alla vendite;
- pricing allo scopo del mantenimento dello status quo [8].

La decisione di sviluppare uno o più di questi obiettivi dipende dai focus dell'azienda. Risulta, quindi, necessario comprendere se questa è focalizzata sugli aspetti competitivi derivanti dal mercato oppure sui costi. Sono questi i punti di partenza sulla base dei quali sviluppare le politiche di prezzo.

Applicare la teoria della “discriminazione dei prezzi” nel suo complesso, come appena definita, non è sufficiente poiché si dovrà inoltre conoscere la sensibilità al prezzo dei diversi segmenti di mercato determinandone i comportamenti, i loro bisogni, le loro aspettative e la loro reale disponibilità di spesa e soprattutto bisognerà porre in essere una effettiva differenziazione di prodotto/servizio che giustifichi la discriminazione stessa.

A tal proposito, occorre considerare che più segmenti di clientela possono trovarsi contemporaneamente presenti all'interno del sistema di erogazione. L'utilizzo della variabile prezzo quale elemento discriminante posto alla base dell'accettazione o del diniego di una richiesta d'acquisto, può generare una sostanziale incompatibilità sia tra soggetti appartenenti a segmenti diversi che tra soggetti appartenenti allo stesso segmento. Vista però l'impossibilità fisica ad adottare questa soluzione la logica alternativa è la ricerca di una giusta combinazione, di un equo bilanciamento. Quando più soggetti si trovano a condividere gli stessi spazi è molto probabile che dialoghino tra loro. Il confronto implica una reale

perdita di immagine aziendale che disattende l'obiettivo di fondo, ossia la *customer satisfaction*⁴ ed ogni sua implicazione operativa. Il cliente "discriminato", ossia il cliente che, per il medesimo servizio offerto anche all'altro cliente, paga di più si sentirà ingannato e truffato dall'impresa e, con ragionevole probabilità, non sceglierà più quella stessa azienda quando vorrà usufruire nuovamente di quel servizio. La soluzione ottimale a questo inconveniente, che può tramutarsi facilmente in un vero e proprio problema aziendale, va ricercata nell'adozione di un sistema di vincoli e di potenziali restrizioni che, agli occhi del consumatore, giustifichino il diverso trattamento offerto.

Questo aggiunge contenuto logico alle variazioni tariffarie e consente di indirizzare efficacemente gli sforzi di vendita verso un predefinito segmento.

Per quanto esposto finora è possibile concludere che l'applicazione di una sola o di poche rigide tariffe è deleterio per l'impresa, di qualunque genere essa sia.

Per quanto illustrato è possibile affermare che a differenti segmenti di clientela dovranno corrispondere differenti tariffe ed anche all'interno di uno stesso segmento sarebbero auspicabili tariffe differenti, tutto ciò sempre in coerenza con l'obiettivo di massimizzazione del profitto, proprio dello yield management.

4.4.2 Previsione della domanda

I principali autori individuano nella previsione della domanda la seconda fase fondamentale del processo di yield management, in cui l'integrazione con la segmentazione analizzata in precedenza è sequenziale [9].

Lo yield management presuppone una domanda variabile, regolabile da leggi probabilistiche, che deve essere gestita attraverso la manovra dei prezzi e delle diverse alternative di marketing. Alla base del buon funzionamento di un sistema di yield è necessaria, quindi, la presenza di un modello capace di prevedere l'andamento della domanda. Si devono effettuare stime separate, sufficientemente puntuali, per ogni segmento, in quanto, diversi sono i comportamenti d'acquisto e non uniformabili sono le variabili decisionali che li influenzano. Il tutto implica l'adozione di politiche sia di prezzo, che di marketing più largamente intese, non standardizzabili.

Il modello di previsione, qualunque sia il contesto di applicazione, si differenzia dagli schemi "tradizionali".

Per quanto concerne le strutture turistico ricettive, ad esempio, ci si basa su due variabili temporali:

⁴ Con questo termine si intende la "soddisfazione del cliente" per il servizio ricevuto e può essere definita come differenza tra qualità percepita e qualità attesa.

- un tempo x precedente la data di fruizione del servizio ed il numero di prenotazioni pervenute fino a quella data;
- le prenotazioni associabili ad ogni singolo giorno.

Il monitoraggio delle prenotazioni giornaliere permette di avere informazioni sull'andamento temporale con cui queste sono pervenute. Tale andamento viene rappresentato dalle curve di prenotazione [10] che sono puntualmente definite per i giorni trascorsi e frutto di approssimazione per gli orizzonti temporali futuri. La sequenza cronologica delle prenotazioni registrate rappresenta invece l'andamento storico della domanda. Questo parametro può essere agevolmente desunto dai dati raccolti nelle giornate che precedono un preciso istante temporale. La disponibilità di entrambe queste serie ha portato a ritenere che si potessero implementare modelli di previsione più accurati da una loro trattazione congiunta.

Nell'ambito dell'approccio combinato i modelli utilizzati possono, a loro volta, basarsi sulla combinazione di strumenti tradizionali dell'analisi di previsione dando luogo a modelli definiti "ibridi". Si possono, ad esempio, applicare modelli di analisi delle serie storiche alle serie temporali delle prenotazioni. Schwartz ed Hiemstra, a tal proposito, hanno sperimentato un modello AR⁵, lo studioso Lee, invece, ha proposto un modello ARMA⁶ multivariato, mentre, infine, Kimes ha suggerito l'utilizzo di modelli a medie mobili, semplici o ponderate [11].

Ciò che effettivamente interessa il management non è però il numero di prenotazioni, anche se questo è il primo dato che muove il modello decisionale ma il numero effettivo di unità di servizio vendute.

L'operatività dello yield management non può pertanto basarsi unicamente sulla previsione delle prenotazioni ma è opportuno che consideri anche una serie di eventualità, non facilmente prevedibili che, nel complesso, si possono definire come "variabili distorsive".

La presenza di queste variabili che, peraltro, sfuggono a qualsiasi tipo di ipotesi previsionale, è la causa della mancata coincidenza tra il numero di richieste ricevute ed il numero di unità realmente acquistate dal consumatore, ed è quindi fonte di numerosi problemi aziendali.

Una corretta gestione ed allocazione della capacità produttiva, quindi, deve, considerare per il futuro, in qualche modo, anche ipotesi difficilmente prevedibili, mentre, per quanto concerne il passato ed il presente le "variabili distorsive" possono essere valutate in maniera concreta tramite gli storici.

Un'elencazione delle più frequenti e rilevanti "variabili distorsive", alcune delle quali riferibili solo ad alcune tipologie di aziende di servizi e, in ogni caso, non esaustiva è la seguente:

⁵ Un modello autoregressivo (AR) considera la generica uscita all'istante k come combinazione lineare delle uscite di tutti gli istanti precedenti a k .

⁶ Un modello autoregressivo a media mobile (ARMA) considera la generica uscita k come combinazione lineare solo di un certo numero, fisso, di uscite precedenti, aggiornate di volta in volta.

- a) **Dow Day of Week**, ovvero il tempo di richiesta del servizio;
- b) **Seasonal Pattern** cioè l'incidenza stagionale delle prenotazioni;
- c) **Prenotazioni on the book** ossia la qualità e l'affidabilità delle prenotazioni;
- d) **Special events impact** ossia l'impatto registrato in concomitanza di alcuni eventi speciali, sia in termini di occupazione che economici;
- e) **Extended Day/Early Departure (Check-out anticipato/posticipato)** cioè la frequenza con la quale si sono verificati soggiorni prolungati o partenze anticipate non previsti;
- f) **No-show** ossia la frequenza con la quale si è verificato che alla prenotazione di un servizio non è corrisposta poi la fruizione dello stesso, in altre parole l'incidenza dei mancati arrivi;
- g) **Denials** ossia il numero di cancellazioni improvvise delle prenotazioni precedentemente effettuate;
- h) **Demand trend** cioè la tendenza della domanda in generale;
- i) **Indici notevoli propri e della concorrenza** quali ad esempio tasso di occupazione, ricavo medio per segmenti di clientela, etc;
- j) **Walk-in** cioè il numero di clienti "di passaggio" che sono stati serviti pur non avendo alcuna prenotazione;
- k) **Last minute booking** ossia il numero di vendite effettuate in prossimità del termine del tempo utile;
- l) **Sell-up** cioè il numero di volte in cui i clienti hanno deciso di pagare ad una tariffa più elevata il servizio non avendo trovato disponibilità all'interno della tariffa scelta in partenza;
- m) **Recapture** ossia, il numero di clienti che, in caso di indisponibilità di risorse da parte dell'azienda, vengono serviti da un'altra azienda della medesima compagnia.

Per una corretta previsione degli andamenti futuri che tenga in considerazione tutte queste variabili sono possibili due ipotesi alternative. Una prima, certamente più complessa sotto il profilo tecnico, ma più puntuale in termini di risultati ottenuti, si fonda sulla previsione separata di tutte le "variabili distorsive" in precedenza menzionate e di altre ancora e permette l'implementazione di accorte politiche di overbooking. Le metodologie di previsione non potranno essere definite a priori in maniera univoca, ma di fronte ad ogni caso concreto si adatteranno sia alle variabili oggetto di scelta che al contesto in cui si inseriscono. In questo modo l'andamento futuro dei volumi di vendita sarà ipotizzato considerando congiuntamente l'andamento delle prenotazioni e le "variabili distorsive".

Un approccio alternativo esula dalle prenotazioni ricevute e formula le sue ipotesi sul numero di unità di produzione effettivamente vendute. La semplicità tecnica ed operativa di questa alternativa si scontra,

come è agevole comprendere, con la minor utilità/efficacia gestionale. In questo modo, inoltre, non è possibile implementare le politiche di overbooking che sono invece un punto nodale dello yield management.

In ottica yield un ulteriore aspetto abbastanza complesso da valutare è la stima dell'andamento della domanda non soggetta a vincoli, intendendo con tale espressione quella parte di clientela che sarebbe possibile soddisfare se la struttura produttiva avesse capacità illimitata. Si è infatti più volte sottolineato che la rigidità della struttura produttiva è una peculiarità di questa tipologia di aziende. Posto che uno degli obiettivi dell'intero processo decisionale è inserire dei limiti alle prenotazioni da accettare sarebbe necessario disporre di una stima accorta della domanda effettiva e non soltanto di quella osservata. In alcuni settori, la sistematica registrazione delle richieste non tradottesi in prenotazioni, può costituire il modo più diretto per stimare la domanda latente che sommata alle prenotazioni consentirebbe di determinare la domanda totale non soggetta a vincoli. In concreto però la registrazione di tali richieste non è così diffusa ed inoltre non tutte le richieste che non sfociano in prenotazioni costituiscono domanda latente: non possono rientrare in questa categoria i declini diretti dei potenziali clienti ma soltanto i rifiuti da parte della struttura produttiva.

Sia per quanto riguarda il settore ricettivo, che il trasporto aereo, sono state individuate metodologie statistiche tese alla stima della domanda totale non vincolata. Infine è necessario ricordare come il processo di erogazione di queste tipologie di servizi turistici sia composto da più fasi, infatti, si può fare riferimento alle diverse possibili tratte che compongono un biglietto aereo acquistato da un unico soggetto per un particolare spostamento e, allo stesso modo, un'unità di servizio nel campo della ricettività turistica può essere acquistata per un arco temporale composto da più giorni. Proprio per questo motivo è necessario che il sistema di yield fornisca un modello capace di considerare, in maniera univoca, l'intera erogazione e non ogni singola fase.

Dopo aver spiegato tutti i fattori che influenzano la domanda si valuta ora, in modo pratico, come si effettua la sua previsione.

La chiave per avere un yield management efficace e di successo è avere una buona conoscenza di quello che sta accadendo nel mercato, di cosa è accaduto al fine di prevedere che cosa accadrà ed effettivamente che cosa può essere fatto per evitare di commettere errori.

Nell'analisi della domanda nel suo complesso si parte quindi da un'analisi della situazione attuale. E' necessario conoscere in maniera approfondita la situazione delle prenotazioni in essere. In questa prima fase pertanto si analizza il numero di prenotazioni ricevute e quante fra queste risultano confermate. Si considerano poi le diverse

fasce comportamentali, l'attendibilità delle prenotazioni stesse ed i programmi che sono stati attuati dall'ufficio vendite a livello promozionale e commerciale.

Nel momento in cui si ha una chiara visione della situazione attuale è possibile procedere con un'analisi del passato. L'analisi del passato è intimamente legata all'analisi della domanda pregressa. L'analisi della domanda pregressa nota anche come *backward demand analysis* considera quanto è accaduto precedentemente tramite gli archivi e le serie storiche. In particolare, per effettuare l'analisi nel migliore dei modi si dovrebbero, come minimo, considerare i tre anni precedenti a quello in corso. Qualora la struttura fosse nuova il monitoraggio non sarà sullo storico delle prenotazioni bensì sarà tarato il più accuratamente possibile sulla concorrenza e sul proprio budget previsionale.

Anche i concorrenti, inoltre, possono rivelarsi una grande fonte di informazione utile per misurare le nostre prestazioni e il trend del mercato; a tale proposito, vi sono molti metodi per scoprire dove questi primeggiano e dove invece sono carenti.

Una volta conclusa la raccolta delle informazioni, si dovrebbe avere un quadro ragionevolmente chiaro circa i livelli della domanda prevedibile.

Si passa, infine, all'ultima fase che certamente risulta essere la più complessa: la stima della domanda futura ossia la *forward demand analysis*.

Quest'analisi è collegata agli eventi previsti di carattere culturale, storico, sportivo che potrebbero influenzare la richiesta di domanda del servizio. Bisogna inoltre considerare le vacanze scolastiche, i ponti e tutti quegli ulteriori accadimenti che potrebbero far variare l'occupazione prevista come azioni promozionali, marketing e *pick-up* ossia erosione di quote di mercato della/dalla concorrenza.

In generale si può affermare che il ricevimento delle prenotazioni, in base all'integrazione della *backward demand analysis* e della *forward demand analysis*, dovrà sempre costantemente monitorare sia i volumi che le qualità stesse: in siffatto modo oltre a conoscere le peculiarità e le abitudini della clientela si riuscirà a prevedere, in modo meno empirico possibile, ogni variazione del trend generale della domanda e a serbare posto per essa.

Un concetto che, dopo aver precisato le nozioni di *pricing* e di domanda, è possibile chiarire e specificare è quello dell'elasticità della domanda al prezzo.

La domanda, infatti, può essere definita come elastica oppure anelastica [12].

Con il termine **domanda elastica** si intende quella domanda che non è turbata da casi di forza maggiore o da eventi esterni, cioè quella domanda che risulta essere particolarmente sensibile alle variazioni di prezzo. Il cliente che effettua un viaggio di piacere (c.d. Cliente *Leisure*) ne è il classico esempio. Questi infatti, qualora dovesse prendere un volo potrebbe partire anche il giorno successivo quello prefissato oppure cambiare la compagnia di volo senza alcun problema, qualora le condizioni di prezzo praticate

dall'azienda di servizi da lui prescelta fossero svantaggiose. Questa tipologia di cliente conferisce un'importanza maggiore al prezzo, successivamente considera il servizio offerto e le sue prestazioni accessorie.

La **domanda anelastica** invece è quella domanda che si presenta in concomitanza di eventi esterni ed è poco sensibile, ed in alcuni casi addirittura insensibile, alle variazioni di prezzo. Il cliente che prende l'aereo per lavoro (c.d. Cliente *Business*) ne è un esempio. Questi infatti è del tutto indifferente alla tariffa offerta, anzi, è disposto a pagare qualsiasi cifra per ottenere un posto sul volo di suo interesse; la priorità di questa tipologia di cliente è infatti quella di ottenere il servizio, a qualunque prezzo.

Se ne desume che la strategia di pricing nei due tipi di domande è differente; ridurre le tariffe e scontare la domanda anelastica non avrebbe alcun senso: essa ci sarebbe comunque, indifferentemente dal prezzo applicato [13]; al contrario offrire tariffe vantaggiose ai clienti con domanda anelastica potrebbe stimolare ed incrementare la richiesta del servizio.

4.4.3 Allocazione della capacità

La capacità è una caratteristica particolarmente critica delle imprese oggetto di questa trattazione, in quanto questa è fissa e limitata, e pertanto necessita di una gestione quanto più accurata e precisa possibile al fine di ottimizzarla.

Il fattore critico nella gestione della capacità è rappresentato dalla possibilità di accettare o meno una prenotazione a prezzo scontato da parte di un potenziale cliente assumendosi il rischio di dover rinunciare ad una probabile vendita futura a prezzo pieno nell'arco temporale che separa il momento in cui si espleta la prenotazione da quello di effettiva fruizione del servizio.

La possibilità di vendere i prodotti/servizi anticipatamente è sicuramente una leva positiva ed un elemento vantaggioso ma i rischi connessi agli elementi di aleatorietà ne impongono una gestione quanto mai accurata. Diviene necessario dotarsi di tutti gli strumenti atti a contenere detto grado di rischio che, ovviamente, varia a seconda della situazione in cui l'azienda si trova.

Di seguito si profilano alcuni degli scenari possibili:

• Disponibilità di servizi:

1. Rischio di mancata vendita: la decisione di imporre un diniego alla prenotazione implica la previsione di una vendita in un momento successivo ad una tariffa più elevata. Nell'ipotesi in cui, detta previsione si riveli troppo ottimistica e non trovi oggettiva realizzazione, il rischio reale è quello di trovarsi con prodotti

invenduti. Si ricordi che, per le aziende oggetto di studio, una mancata vendita corrisponde ad una perdita in quanto non è possibile trasferire i servizi nel tempo.

2. Rischio di mancati ricavi: nell'ipotesi contraria, si decide di accettare la prenotazione supponendo di non ricevere, in momenti successivi, richieste d'acquisto da parte di clienti disposti a pagare un prezzo più elevato. Anche in questo caso, se lo scenario prefigurato non trova realizzazione il rischio effettivo è rappresentato da una perdita di componenti positivi di reddito.

• **Indisponibilità dei servizi:**

1. Rischio di mancata vendita: un'unità di servizio per la quale si era già ricevuta una prenotazione può essere considerata indisponibile e quindi si tenderà a rifiutare qualsiasi altra prenotazione. In questo modo si rischia di incappare in una mancata vendita, cui necessariamente consegue una diminuzione di ricavi nell'ipotesi in cui si verificano situazioni di *no-show* o cancellazioni.

2. Rischio connesso all'overbooking: in questo caso si prevede che non tutte le prenotazioni ricevute si trasformino in effettivi atti d'acquisto a causa appunto di possibili *no-show* o cancellazioni e si accetta un numero di prenotazioni superiori alle effettive capacità di offerta disponibili. In questo modo ci si espone al rischio di *oversale*.

La gestione/allocazione della capacità produttiva si sostanzia pertanto sia nella definizione di un livello di prenotazioni superiore alla capacità di offerta (overbooking) atta a compensare possibili mancate conclusioni dei processi di acquisto sia nell'allocazione dei contingenti ai vari segmenti tariffari, in maniera tale da rendere possibile la massimizzazione dei ricavi. Riguardo quest'ultima decisione occorrerà scegliere sia il numero di unità da vendere a prezzo scontato che l'arco temporale durante il quale lasciare aperta detta opportunità di acquisto. Queste scelte rispondono all'obiettivo del raggiungimento della massima occupazione ma potrebbero entrare in collisione con la massimizzazione dei ricavi. Il fine cui l'azienda tende è la vendita del maggior numero di servizi al prezzo unitario più elevato. Ad esempio, a tal proposito è bene precisare che qualora le camere all'interno della struttura ricettiva, o i posti su di un aeromobile allocati ai segmenti tariffari più elevati fossero esauriti non si dovrà rifiutare nessuna ulteriore richiesta proveniente da detti segmenti se rimangono ancora unità realmente disponibili anche se allocate a segmenti inferiori; questo sempre coerentemente all'ottica di massimizzazione del profitto.

Spesso i segmenti di clientela meno sensibili al prezzo, ai quali quindi è possibile applicare tariffe più elevate, sono quelli che prenotano con un contenuto margine di anticipo o non prenotano affatto. Quanto appena affermato è confermato dalla figura seguente (Fig. 4.1) che rappresenta la richiesta di vendita nel tempo nel caso di clienti sensibili al prezzo e nel caso di clienti non sensibili a quest'ultimo. La figura mostra che mentre i clienti attenti al prezzo tendono a prenotare con un ampio anticipo rispetto alla data di

fruizione del prodotto/servizio per avvalersi di eventuali riduzioni della tariffa base, i clienti indifferenti al prezzo, o comunque meno sensibili a questa variabile, tendono ad acquistare il prodotto/servizio in una data molto prossima al suo effettivo utilizzo e sono, invece, molto più sensibili ed attenti ad altre tipologie di variabili, come ad esempio la qualità del prodotto/servizio base ed i servizi accessori offerti insieme ad esso.

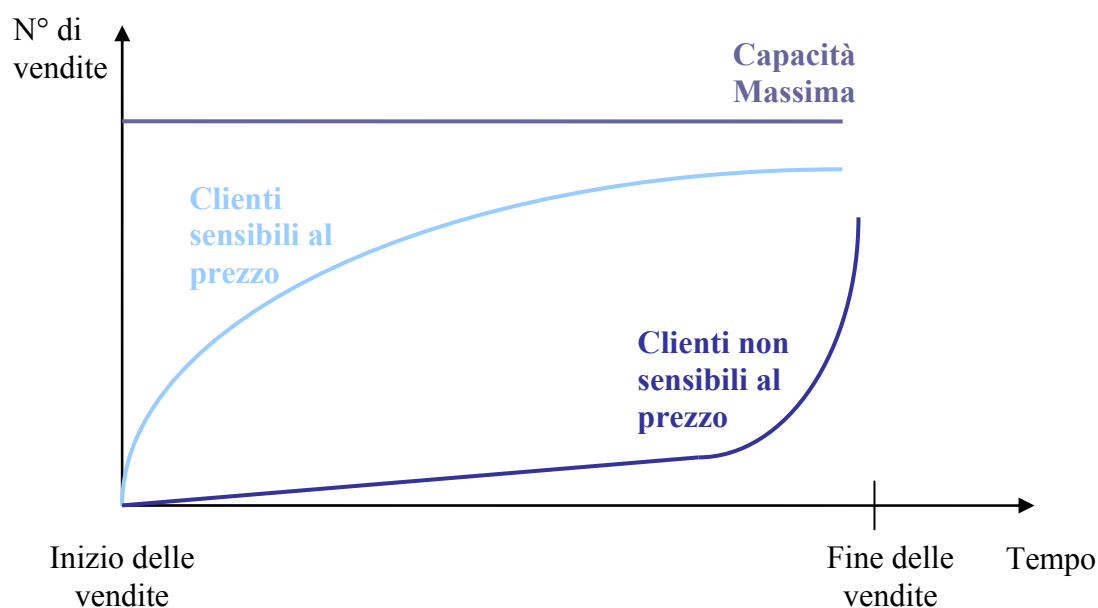


Figura 4.1: Andamento delle vendite per i diversi segmenti di clientela.

Un altro fattore molto importante per la gestione della capacità è il tempo: lo yield management per risultare efficace, deve attuare controlli progettati e eseguiti con un certo anticipo nel tempo. Come rappresentato in figura 4.2 l'entità del controllo varia in relazione al tempo in cui questo viene attuato.

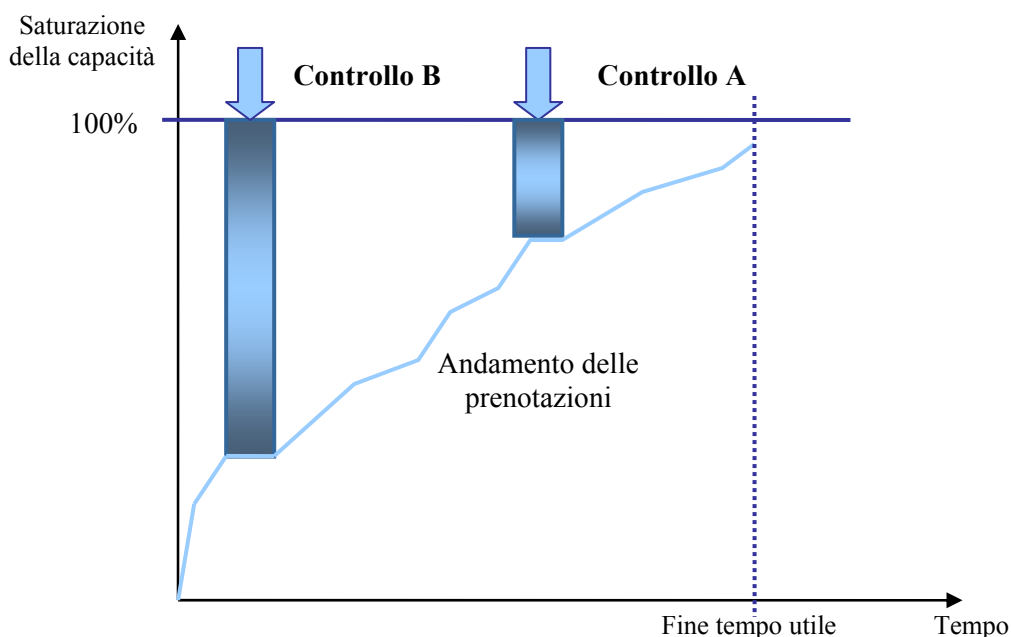


Figura 4.2: Incidenza del controllo sulla capacità

Analizzando il grafico si può notare che il controllo A è proiettato nel futuro solo di alcuni giorni, in prossimità del giorno di fruizione del servizio, quando il livello delle prenotazioni è già molto alto, per cui si può tentare di vendere nelle condizioni ottimali soltanto una piccola parte della capacità, cioè quella residua. Viceversa, il controllo B è proiettato in un futuro più lontano, nel momento in cui il livello delle prenotazioni è ancora relativamente basso, pertanto si può tentare di vendere a tariffe più vantaggiose per l'azienda una grande parte della capacità totale di servizio, che risulta essere ancora disponibile [14].

4.4.4 Prenotazione e vendita

La vendita è l'elemento finale del sistema di yield management. Tutte le ipotesi elaborate nelle fasi precedenti devono trovare riscontro in un incremento della componente positivo di reddito.

L'elemento caratteristico della vendita nelle imprese oggetto di analisi è la possibilità di vendita anticipata del prodotto/servizio: la prenotazione. Proprio per questo, il processo di prenotazione diventa, di fatto, l'elemento chiave della fase di vendita e nelle realtà più complesse implica la presenza di un adeguato sistema informatico.

Inoltre, una parte rilevante delle vendite di un'impresa che opera in campo turistico è contraddistinta da rapporti duraturi, di lungo periodo, siglati da contratti formalizzati con ampio margine di anticipo ed in maniera consuetudinaria, che esulano dal campo yield. Si pensi, a titolo di esempio, ai rapporti che

intercorrono tra un'azienda turistico-ricettiva ed un tour operator o, a quelli stipulati tra una compagnia aerea ed alcuni enti no profit o clienti frequenti. Per questa ragione un'azienda che volesse impiegare sistematicamente le tecniche dello yield management e considerare, al contempo, i ricavi complessivi deve integrare dette tecniche con le politiche di vendita sul medio e lungo periodo.

A questo punto è possibile delineare le caratteristiche di un sistema di prenotazione da implementarsi in ottica yield. Le considerazioni che seguono saranno di inquadramento generale, in quanto ogni settore ha le proprie caratteristiche ed una generalizzazione risulterebbe fuorviante.

Il primo elemento oggetto di analisi è la struttura dell'offerta. Il sistema deve consentire una strutturazione quanto mai flessibile. La complessità e la variabilità dei mercati di riferimento impone al management di predisporre possibilità di acquisto flessibili e modificabili tempestivamente. L'operatività dello yield management è correlata poi sia alla gestione della disponibilità fisica delle unità di servizio, che alla relativa tariffazione. La prenotazione, nonché la conseguente operazione di vendita, sono poste in funzione non soltanto della disponibilità del "supporto fisico" ma anche delle decisioni di allocazione. La disponibilità, in prima istanza, è tariffaria. Quello che si offre al cliente che formalizza una prenotazione è una classe di prezzo cui è associato un servizio. Tant'è che, per la medesima unità di offerta, in momenti diversi possono essere richiesti prezzi diversi in funzione del momento in cui viene effettuata la prenotazione e dell'andamento della domanda, della sua stagionalità. Quest'ultima, in un sistema di yield, non è predefinita e fissa ma è correlata all'effettivo andamento delle vendite. I limiti temporali degli archi stagionali sono mobili e non si può stabilire preventivamente la loro durata.

I modelli di domanda sui quali è stata ipotizzata la politica di yield management devono essere costantemente controllati. A detto proposito un ruolo chiave è giocato dal personale di front-line che entra costantemente in contatto con la clientela. Spesso è lo stesso sistema di prenotazione che facilita e supporta questo interscambio. Tra gli aspetti positivi di questa fase vi è sicuramente la possibilità di percepire tempestivamente il grado di soddisfazione del cliente cogliendo eventuali scostamenti negativi e di apportare in modo repentino le opportune modifiche.

Ad un sistema di prenotazione tradizionale sono richieste funzioni assai elementari come gestire un'anagrafica clienti e permettere la effettiva realizzazione della prenotazione. Ciò che rileva in ottica yield è invece qualcosa di più: le specifiche che devono essere fornite riguardano sia i contatti che i referenti anagrafici che le cancellazioni. I contatti sono quelle semplici richieste di informazioni che non danno luogo a prenotazione, sono interazioni tra potenziali clienti e personale di contatto che rappresentano però dati importanti per valutare quantitativamente e qualitativamente la domanda effettiva che si è rivolta all'azienda. Una prenotazione che non si trasforma in domanda effettiva, sia per un rifiuto

da parte del cliente, che per un diniego da parte dell'azienda contribuisce comunque a formare la domanda effettiva. Diviene così cruciale il ruolo giocato dal personale di front-line che deve essere capace di cogliere sia le caratteristiche della domanda potenziale che le motivazioni che hanno portato ad un rifiuto, dati questi che solitamente sono contenuti proprio nella scheda di prenotazione. Una cultura aziendale orientata allo yield non può prescindere da queste sensibilizzazioni ed il management deve approntare gli strumenti necessari affinché possano realizzarsi.

Solamente i dati relativi ai referenti anagrafici, specie quando i soggetti che eseguono la prenotazione non coincidono con quelli che poi materialmente usufruiranno del servizio, talvolta vengono rilevati in maniera incompleta.

A riguardo delle cancellazioni il sistema deve permettere di rilevare la data ed il motivo. Se si conosce il momento in cui la cancellazione è effettuata sarà più agevole, in futuro, modellare il sistema di allocazione. Allo stesso modo la motivazione che ha determinato la mancata conclusione del contratto di acquisto sarà un utile supporto informativo nel processo di segmentazione.

Le osservazioni sulla domanda che un sistema di yield impone non si esauriscono con l'affinamento del sistema di prenotazione. Altri dati, raccolti in momenti diversi sono indispensabili al buon funzionamento del processo. Si fa riferimento in modo particolare a no-show, oversale e show-up.

La mancata presentazione di un cliente al check-in, a fronte di una prenotazione non permette di comprendere se questa è effettivamente andata a buon fine, così come per valutare la bontà e la significatività delle politiche di overbooking deve essere possibile un'attenta gestione degli oversale.

Infine gli eventuali contatti realizzati con clienti walk-in non sono annoverati in fase di prenotazione, ma presso il *front desk*. Anche questi però contribuiscono a formare la domanda che è possibile/opportuno soddisfare. All'interno del sistema quindi deve essere inserito un espediente capace di monitorare queste informazioni.

4.5 Analisi dello Yield management

4.5.1 Vantaggi

All'implementazione di un sistema di yield management seguono una serie di benefici diretti ed indiretti.

Il beneficio diretto è costituito dall'ottimizzazione dei ricavi ottenuta attraverso una sistematica attività di ricerca e di cattura dell'intera domanda potenziale che l'azienda è in grado di soddisfare, minimizzando il rischio di mancati ricavi in relazione al posizionamento corrente e individuando contemporaneamente nuovi segmenti di domanda da servire. Tale beneficio ha conseguenza diretta sul conto economico dell'impresa [3].

I benefici indiretti sono molteplici: incremento dell'attenzione al cliente, maggiore correlazione tra strategia e processi operativi aziendali, con conseguente incremento delle capacità di pianificazione e controllo, razionalizzazione gestionale delle risorse, incremento del patrimonio informativo aziendale e della consapevolezza della sua rilevanza nonché delle tecnologie ad esso associate.

Il principale vantaggio dello yield management è quello di stabilire una strategia di prezzo efficiente e dinamica fatta “su misura” per ogni singolo cliente. Il concetto base da cui parte l’implementazione di questa tecnica gestionale è che il mercato è soggetto a continui mutamenti, così come il processo d’acquisto dei consumatori, che risulta essere influenzato da numerose variabili aleatorie. Il management aziendale deve essere dinamico e seguire in modo repentino i mutamenti del mercato, per fare questo deve poter essere messo in condizione di variare velocemente le scelte compiute per non disattendere le aspettative dei clienti. Non è pensabile definire una strategia di prezzo destinata a rimanere invariata all’interno di un definito intervallo temporale, ma bisogna costantemente monitorare il mercato e recepire anche i più impercettibili segnali che esso manda. Da quanto detto è possibile concludere affermando che il primo vantaggio dello yield management è quello di essere dinamico e quindi effettivamente al passo con le esigenze dei clienti. Dalla conoscenza dei clienti e delle loro esigenze però lo yield ne trae un vantaggio personale, che rappresenta la sua chiave di successo. Questa tecnica riesce infatti a fornire ad ogni cliente esattamente ciò che vuole in termini di servizio offerto ma soprattutto di prezzo. Con un’attenta politica di segmentazione della clientela e successivamente di pricing lo yield riesce a conseguire il proprio obiettivo, che poi è caratteristico di tutte le aziende che vogliono sopravvivere nel mercato: la massimizzazione del profitto.

Attraverso le diverse fasce di prezzo che l’impresa applica riesce ad attirare a sé anche i clienti che a tariffa piena non avrebbero mai usufruito del servizio, saturando così la propria capacità produttiva. La situazione descritta è rappresentata nella figura 4.3.

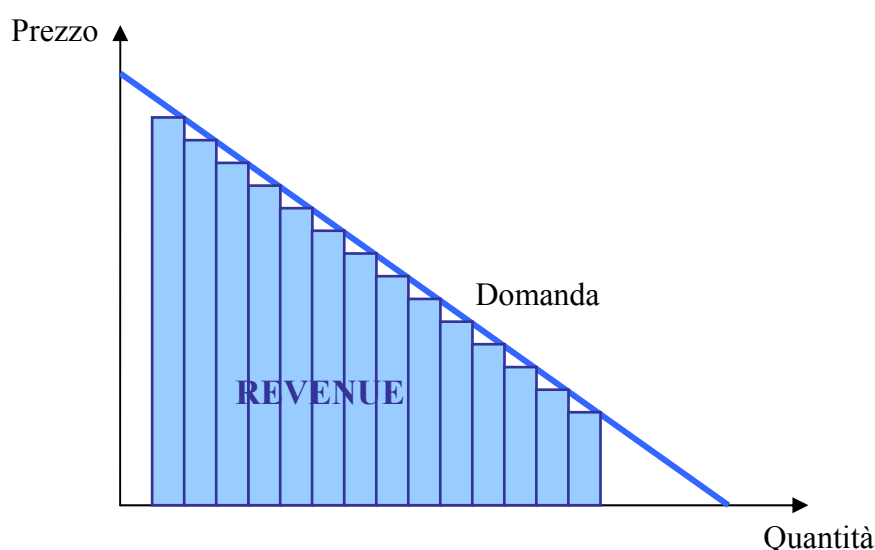


Figura 4.3: Revenue ottenuto con una politica di discriminazione dei prezzi

In una politica di discriminazione dei prezzi ogni rettangolo rappresenta il profitto conseguito con una certa tariffa mentre la somma di tutti i rettangoli rappresenta il profitto totale ottenuto; l'area ottenuta in questo caso è molto superiore di quella ottenibile utilizzando un'unica tariffa alla quale cui corrisponde un unico rettangolo.

Il principale vantaggio dello yield management, che ha permesso un'enorme diffusione in meno di venti anni, è che *[..si può pensare ad esso come ad un gioco in cui vincono tutti i partecipanti..]* [14].

In primo luogo si riescono a garantire profitti all'impresa, ma allo stesso tempo si riescono a soddisfare tutte le tipologie di consumatori: il cliente con poca disponibilità economica potrà usufruire di un servizio altrimenti inaccessibile tramite offerte promozionali o last-minute, il cliente business riuscirà molto probabilmente ad usufruire del servizio anche con un preavviso molto breve ed infine i clienti "affezionati", cioè i clienti che garantiscono all'azienda profitti distribuiti e costanti durante tutto l'orizzonte temporale di riferimento avranno sempre la garanzia di un posto riservato.

Lo yield management non è quindi, come si potrebbe pensare, un approccio sistematico per abusare della pazienza dei clienti, ma una tecnica che cerca di allineare il pricing delle strutture ricettive con la domanda che in quel momento esiste nel mercato. Infatti, il concetto fondamentale legato a questa tecnica è che una tariffa dipende in primo luogo dalla domanda specifica per quella data. Per il medesimo servizio potranno essere richieste tariffe totalmente differenti a seconda della domanda per quel periodo, della stagionalità. Ad esempio il concetto di "alta stagione" e di "bassa stagione" per la tariffa di una camera in un albergo segue esattamente questa logica, nello specifico si considera l'alta stagione come affetta da domanda anelastica in quanto anche applicando tariffe elevate, essendo altrettanto elevata la domanda del servizio, si riuscirà in ogni caso a saturare la capacità e ad ottenere il massimo profitto, al contrario la bassa stagione si considera affetta da domanda elastica e, per scongiurare il rischio di non saturare la capacità, essendo la domanda per il servizio bassa, si cerca di attirarla a sé proponendo prezzi bassi.

4.5.2 Pregiudizi e critiche

Le tecniche di yield sono nate in un contesto operativo per risolvere in modo pratico problemi che si presentano quotidianamente alle aziende, questo ha comportato una grande frammentazione sia sul lato teorico, di inquadramento sistematico delle metodologie, che operativo, dove molti si comportavano proprio secondo le tecniche senza aver compreso a fondo il loro significato.

Questa situazione ha favorito l'insorgere di una serie di pregiudizi sullo yield management che hanno poi condizionato la sua diffusione.

Il primo e più grosso pregiudizio è quello di credere che lo yield sia solamente un sistema computerizzato. Indubbiamente la tecnologia informatica ricopre un ruolo molto importante nel sistema di yield management: in primo luogo perché il trattamento automatico dei dati ci permette di utilizzare, analizzare e creare una grande quantità di informazioni in breve tempo, nonché di razionalizzare i vari processi per il calcolo del prezzo; la seconda ragione è che l'industria informatica del software, in special modo negli Stati Uniti, ha creato dei programmi "*yield management oriented*" non solo per grandi catene o strutture, ma anche per piccoli alberghi; così facendo, le case di software hanno permesso anche alla maggioranza degli imprenditori del settore ricettivo di conoscere questa tecnica. Il legame informatica-yield management, però, non ha portato solo benefici, in quanto, come detto, molti albergatori hanno erroneamente creduto e credono ancora che lo yield management possa essere applicato semplicemente possedendo un computer ed un software specifico senza porsi il problema di capire i principi, i meccanismi e le strategie del metodo [15]. Un altro errore frequente, che è stato commesso da vari albergatori al momento dell'acquisto del sistema informatico, è quello di non aver effettuato preventivamente una serie di considerazioni da cui dipende l'efficacia e l'efficienza del sistema stesso, come ad esempio l'analisi e l'adeguamento della "rete di sostegno", dove, con rete di sostegno si intende l'ambiente in cui è calata la tecnologia informatica che, in un'impresa, è costituito dall'insieme delle relazioni organizzative, amministrative e culturali.

Generalmente l'approccio allo yield management viene accomunato ad un sistema computerizzato, in realtà non si tratta né di un sistema computerizzato né tanto meno di un insieme di regole matematiche o statistiche. Si tratta invece di un modo di incrementare il volume di vendita migliorando i servizi per affrontare in maniera più costruttiva e per soddisfare le esigenze del mercato. Nella realizzazione dello yield, il supporto informatico resta comunque di grande importanza ma che non potrà mai sostituire elementi come la cultura aziendale e direzionale orientata allo yield stesso.

Un altro pregiudizio molto diffuso è che questa tecnica funzioni soltanto quando la domanda è superiore all'offerta. In realtà questo pregiudizio nasce poiché la parte del sistema di yield maggiormente conosciuta è quella relativa alle politiche di overbooking e si è così indotti a pensare che il tutto sia utile solo quando ci si trova nella agevole posizione di dover scegliere quale tipo e quanta parte della domanda potenziale dover soddisfare. Detta situazione, però, non rappresenta il quotidiano per numerose realtà e pertanto si ritiene che le tecniche di yield siano uno strumento utile solo ad una piccola élite.

A rappresentare uno degli elementi base del processo di yield non è quindi la politica di gestione dell'overbooking bensì la previsione della domanda.

Il management deve essere messo nella condizione di conoscere quali saranno i periodi di restrizione al fine di stimolare la domanda potenziale. Lo yield aiuta a modificare la struttura tariffaria e a controllare le vendite al fine di ottenere un migliore sfruttamento della capacità produttiva. Il suo fine pertanto non è, come già precisato, quello di gestire picchi di domanda che superano l'offerta potenziale ma quello di favorire un migliore sfruttamento della capacità produttiva.

Un'altra falsa idea è reputare lo yield management semplicemente come una politica di vendita basata sulla riduzione del prezzo nei periodi in cui la domanda è contenuta e conseguente aumento nei periodi in cui la stessa è più elevata.

In realtà invece lo yield si focalizza su quanto è necessario vendere di un determinato prodotto/servizio ad un determinato prezzo in un preciso istante temporale; non suggerisce cambiamenti tariffari ma indica quando aprire e chiudere determinate classi di prezzo attraverso il supporto informativo fornito dal monitoraggio continuo dell'andamento della domanda.

Un ultimo pesante pregiudizio sullo yield è che questo sia incompatibile con un buon servizio. In realtà, come già detto precedentemente, questa tecnica è proprio valida in quanto riesce a coniugare le esigenze di diversi soggetti oltre che quelli dell'impresa stessa.

L'unico parametro effettivamente critico per l'erogazione di un buon servizio è rappresentato dall'overbooking.

La critica principale fatta allo yield management è quella di abbassare troppo le tariffe. Le tariffe non vanno abbassate eccessivamente in quanto il cliente potrebbe associare al prezzo basso una qualità del servizio altrettanto bassa, che non rispecchia le proprie aspettative e quindi essere indotto a rifiutare il servizio per tale motivo. In realtà è vero che lo yield è una strategia per ottimizzare i profitti dell'azienda ma non deve indurre a perseguire ad ogni costo il tasso d'occupazione del 100% in quanto in tal modo si rischierebbe di ridurre il margine di profitto poiché la clientela che vuole acquistare il servizio, non sarà più disposta a spendere troppo avendo una sensazione di bassa qualità.

Psicologicamente, purtroppo, l'uomo associa ad una tariffa bassa un livello equivalente del servizio, in quanto egli raramente conosce le complesse dinamiche del mercato tra le quali, in particolare, l'elevata concorrenza che spinge le imprese a diminuire, anche di molto, il prezzo di un prodotto rispetto al suo valore reale pur di riuscire a ricavare da esso un minimo profitto ed erodere clienti alle altre imprese del settore.

Un'altra critica che potrebbe essere fatta alla tecnica è la sua difficile implementazione che richiede comunque una serie di cambiamenti di non facile realizzazione tra cui una ristrutturazione organizzativa e delle mansioni, una comunicazione efficace ed un coordinamento elevato, una distinzione chiara dei ruoli e delle relative responsabilità e soprattutto una formazione aziendale continuativa a tutti i livelli. Un cambiamento aziendale, infatti, porta con sé l'obbligo di disapprendere convinzioni, atteggiamenti e valori ed impararne di nuovi, di dimenticare comportamenti fino a quel momento considerati normali e di routine per far posto a nuove procedure. In tale ottica, l'unico strumento a disposizione delle imprese, per non trasformare il cambiamento in un fallimento, è il training del personale.

4.6. La gestione delle risorse

Gli strumenti per una razionale gestione delle risorse, in ogni caso in numero finito, sono sostanzialmente due:

1. La definizione del *Protection Level* ossia la scelta, per ogni segmento di mercato individuato, del numero di unità da vendere a tariffa piena e conseguentemente, per differenza, il numero di unità da vendere a tariffa scontata, e la definizione dell'arco temporale durante il quale lasciare aperta detta opportunità di acquisto;
2. La gestione della politica di *overbooking* cioè, in primo luogo la decisione di utilizzare o meno questa tecnica che, come detto, comporta numerosi vantaggi ma anche diversi rischi che vanno opportunamente bilanciati e, successivamente, nel caso di utilizzo, la definizione del limite di prenotazioni in overbooking da accettare.

Il fine ultimo della gestione della capacità resta comunque il massimo utilizzo delle stesse e la massimizzazione dei ricavi da esse conseguibili.

4.7 Concetto di Protection Level

Lo yield management procede, in primo luogo, attraverso una discriminazione di prezzo che, nel caso più semplice prevede l'applicazione di due tariffe differenti: una tariffa intera T_i ed una tariffa scontata T_s .

Il problema da affrontare consiste, in questo caso, nella decisione quantitativa del numero di risorse da vendere a prezzo scontato e, per differenza, ricavare il numero di quelle da vendere a prezzo intero.

Questo problema, che in prima lettura può sembrare semplice, risulta invece abbastanza complesso in quanto presuppone una approfondita analisi e segmentazione del mercato, per evincere i comportamenti d'acquisto e la sensibilità al prezzo delle diverse tipologie di clienti.

Nel caso in esame si individuano solo due tipologie di clienti:

- Un *cliente Leisure*, che utilizza il servizio offerto dall'impresa soprattutto per svago, molto sensibile al prezzo e con l'abitudine a prenotare con molto anticipo;
- Un *cliente Business*, che utilizza il servizio soprattutto per ragioni lavorative, indifferente o poco sensibile al prezzo e con l'abitudine a prenotare in una data molto prossima a quella di utilizzo del servizio offerto.

A tal proposito si definisce “**Booking Limit**” o “Limite di booking” il numero massimo di risorse (ad esempio camere, posti a sedere su un volo, cabine, etc.) vendibili a prezzo scontato, e “**Protection Level**” o “Limite di protezione” il numero di risorse “protette” e non vendute a prezzo scontato, benché ce ne sia la domanda, in quanto si ritiene di poterle vendere successivamente nel tempo a prezzo intero[16].

$$\text{Booking Limit} = \text{Risorse totali} - \text{Protection Level} \quad (1)$$

L'equazione (1) è mostrata graficamente nella figura seguente.

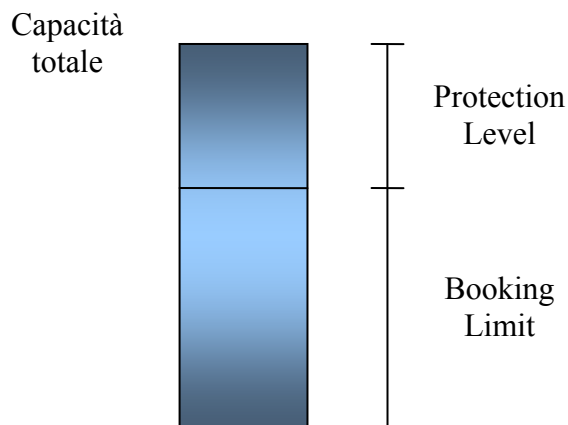


Figura 4.4: Protection Level e Booking Limit

Questo caso è il più semplice esistente in letteratura in quanto si considerano solo due classi di prezzo, una intera ed una scontata, e due soli segmenti di clientela, uno business ed uno leisure.

Nella realtà applicativa, soprattutto nei contesti molto grandi, si presentano invece diverse tariffe, di certo in numero superiore a due, proprio per tentare di saturare al massimo l'elevata capacità produttiva e di assecondare le esigenze e le aspettative delle diverse tipologie di clienti, precedentemente segmentati in diverse classi, accomunati dagli stessi bisogni e dalla stessa disponibilità a pagare.

In questi casi, occorre individuare un protection level relativamente ad ogni segmento di mercato individuato; l'obiettivo dell'ottimizzazione consiste quindi nell'individuare il numero di unità di servizio da proteggere per ogni classe di prezzo, il tutto sempre in modo coerente all'obiettivo di massimizzazione del profitto, proprio della tecnica gestionale dello yield management.

4.7.1 Metodi risolutivi nel caso di due tariffe

La presenza di due sole tariffe di vendita del prodotto/servizio è caratteristica solo di piccoli contesti aziendali e, come già detto, è lontana dai contesti nei quali si trovano ad operare le aziende di servizi di grosse dimensioni.

Queste ultime infatti per saturare la propria capacità produttiva nel miglior modo possibile, cioè quello più profittevole, tendono ad offrire il prodotto/servizio a diverse tariffe in modo da riuscire a catturare quanta più clientela possibile, offrendo ad ogni segmento esattamente ciò che esso vuole, al massimo prezzo che quest'ultimo è disposto a pagare.

Il modo di operare appena illustrato è infatti coerente e fedele all'essenza dello yield management che consiste nel vendere al cliente giusto il servizio giusto al prezzo giusto nel momento giusto il tutto con il fine ultimo di massimizzare il profitto [17].

In ogni caso può risultare comunque utile e vantaggioso, come punto di partenza per la trattazione, considerare le diverse metodologie attraverso le quali è possibile individuare il valore del protection level ottimo, cioè che massimizza il profitto aziendale, nel caso di due sole tariffe per poi estendere le considerazioni anche ai casi più complessi ma più reali che, invece, considerano contemporaneamente la presenza di diverse tariffe offerte per soddisfare i diversi segmenti di clientela.

La logica di base e le riflessioni sul protection level sono infatti le medesime sia nel caso di due tariffe che nel caso di più di due tariffe, ciò che cambia sono solo i metodi di determinazione dello stesso che nel primo caso risultano essere più semplici ed immediati mentre nel secondo caso risultano leggermente più complessi ed impegnativi.

4.7.2 Analisi deterministica

L'analisi deterministica è la più lontana, in termini assoluti, dai contesti operativi reali in quanto presuppone la perfetta conoscenza della domanda delle risorse a tariffa intera, praticamente impossibile, e la presenza di due sole tariffe alle quali è offerto il servizio.

In questo caso è possibile determinare in maniera elementare il protection level come:

$$\text{Protection Level} = \text{Domanda a tariffa intera} \quad (2)$$

Il protection level, in pratica, viene posto pari alla domanda a tariffa intera, che è nota in maniera deterministica e quindi quantificabile ed, in tal modo, attraverso una semplice uguaglianza si determina il valore dell'unica incognita del problema [18].

Nei contesti applicativi reali la domanda non è mai nota in maniera deterministica poiché essa dipende da una moltitudine di fattori, alcuni dei quali non esattamente prevedibili con largo anticipo, altri non precisamente prevedibili in modo assoluto.

Le motivazioni che spingono i singoli consumatori all'acquisto di un servizio, che sommati danno poi origine alla domanda totale di quel servizio, seguono talvolta logiche avulse da qualsiasi standardizzazione e pertanto rendono praticamente impossibile la conoscenza analitica ed esatta delle richieste totali per un servizio.

Per questa ragione accanto a questa tecnica di analisi deterministica se ne considerano altre euristiche e probabilistiche che meglio rispecchiano le limitate conoscenze, tipiche dei contesti aziendali reali, caratterizzate da forte incertezza ed instabilità.

4.7.3 Analisi tramite albero delle decisioni

L'albero delle decisioni consente di visualizzare in modo grafico, compatto e semplificato un'analisi decisionale.

L'analisi decisionale è una modalità di analisi dei problemi o delle decisioni, che consente di individuare, in condizioni d'incertezza, le conseguenze, in questo caso economiche, dell'attuazione di determinate strategie.

L'applicazione dell'analisi decisionale comporta sei fasi:

- Identificare la decisione, includendo la selezione delle opzioni decisionali da studiare. Definire il periodo temporale della decisione (ad esempio i costi e i benefici in un anno, o i costi e i benefici durante tutto l'arco della vita di una persona). Determinare da quale prospettiva la decisione deve essere presa;
- Strutturare la decisione e le conseguenze di ogni opzione decisionale nel tempo;
- Valutare la probabilità di accadimento di ogni conseguenza;
- Determinare il valore degli *outcomes*, in termini di unità monetaria;
- Selezionare l'opzione con il maggiore risultato atteso o il più basso costo atteso, cioè la più conveniente dal punto di vista economico;
- Determinare la robustezza delle decisioni impiegando una analisi di sensibilità e variando i valori di probabilità e i risultati all'interno di un intervallo di valori.

Un albero decisionale, come anticipato, è spesso impiegato per rappresentare un processo decisionale. Le decisioni vengono rappresentate all'interno di un albero decisionale tramite un quadrato (nodo decisionale), i cui i rami e/o bracci rappresentano le opzioni alternative rilevanti. I cerchi indicano i nodi probabilistici, dove gli eventi futuri sono sotto il controllo del decision maker e i cui risultati sono incerti.

Le probabilità di accadimento sono assegnate a ciascun ramo a partire da ogni singolo nodo probabilistico. Le probabilità di accadimento di ogni nodo probabilistico devono chiaramente totalizzare 1. Per ottenere il costo complessivo di ogni nodo decisionale, i valori degli *outcomes* (costi) vengono moltiplicati per le loro rispettive probabilità di accadimento.

I dati per condurre analisi decisionali possono essere ottenuti da molteplici risorse, come ad esempio storici, database ed opinioni di esperti.

Si consideri ora, l'applicazione di questo strumento di analisi alla tecnica dello yield management.

Indicando con:

- T_i la tariffa intera;
- T_s la tariffa scontata;
- D_i la domanda a tariffa intera;
- D_s la domanda a tariffa scontata.

la decisione da prendere consiste nello scegliere se portare o meno il protection level da $(S - 1)$ ad S .

L'albero delle decisioni corrispondente è rappresentato nella figura seguente.

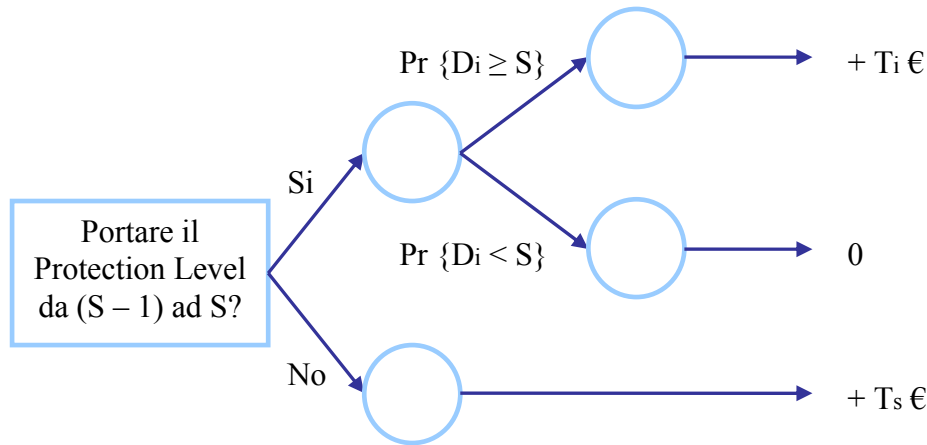


Figura 4.5: Albero decisionale per un problema di yield management

L'albero decisionale rappresenta un "iter logico" coerente con gli obiettivi dello yield e sintetizzabile con una semplice conclusione: risulta conveniente aumentare il protection level dal livello (S-1) al livello S solo se la probabilità che la domanda del servizio a tariffa intera è superiore o uguale al protection level S.

Dal punto di vista matematico, con la stessa logica probabilistica è possibile affermare che è opportuno e conveniente incrementare il numero di risorse riservate alla tariffa piena se si verifica che:

$$T_i * \Pr \{ D_i \geq S \} \geq T_s \quad (3)$$

da cui ricordando che la probabilità ha valore unitario:

$$T_i * [1 - \Pr \{ D_i < S \}] \geq T_s \quad (4)$$

ed infine, isolando la probabilità ad un membro:

$$\Pr \{ D_i < S \} \leq (T_i - T_s) / T_i \quad (5)$$

È possibile concludere affermando che se la probabilità che la domanda a tariffa intera D_i sia minore del protection level S è inferiore o pari ad un certo numero, dipendente dalla tariffa intera e da quella scontata, entrambe note e misurabili allora la scelta di incrementare il protection level fino al valore S è economicamente conveniente.

Infine, per quanto riguarda il processo di analisi decisionale nel suo complesso, sembra opportuno sottolineare che questo fornisce una serie di informazioni aggiuntive alla mera risposta finale: infatti, si evidenziano sia la struttura

sottostante al processo decisionale che le assunzioni di base a fondamento della decisione presa, rendendo quest'ultima più chiara, logica e facilmente comprensibile.

4.7.4 Analisi marginale

Il metodo dell'analisi marginale è correntemente impiegato ogni volta che si ha incertezza sulla domanda e bisogna prendere decisioni in anticipo rispetto al momento in cui si manifesta la domanda stessa.

Indicando con:

- D_i la domanda a tariffa intera;
- C_u il costo unitario sostenuto per aver sottostimato la domanda a tariffa intera;
- C_o il costo unitario sostenuto per aver sovrastimato la domanda a tariffa intera;

questo metodo funziona in modo semplice e logico.

Operativamente si incrementa il numero di risorse destinate ad essere vendute al prezzo intero fino a quando il beneficio atteso sull'ultima unità di risorsa allocata non supera la perdita attesa sull'ultima unità allocata.

In modo compatto cioè occorre verificare:

$\Pr\{\text{Sottostima } D_i\} * \text{Costo sottostima} \geq \Pr\{\text{Sottostima } D_i\} * \text{Costo sottostima}$

che, in termini matematici può essere scritto come :

$$\Pr \{ D_i \geq S \} * C_u \geq \Pr \{ D_i < S \} * C_o \quad (6)$$

da cui, essendo la probabilità totale pari all'unità:

$$[1 - \Pr \{ D_i < S \}] * C_u \geq \Pr \{ D_i < S \} * C_o \quad (7)$$

ed infine svolgendo la parentesi ed isolando i costi si ottiene:

$$\Pr \{ D_i < S \} \leq C_u / (C_u + C_o) \quad (8)$$

Anche in questo caso, quindi il protection level risulta facilmente determinabile una volta individuati il costo sostenuto per aver sottostimato e/o sovrastimato la domanda del servizio a tariffa intera.

Se inoltre si nota che il costo sostenuto per aver sottostimato la domanda a tariffa piena rappresenta un mancato guadagno in quanto si sarebbero potute vendere risorse a tariffa intera anziché, come è stato fatto, a tariffa scontata e, per tale ragione, il costo suddetto risulta esprimibile come differenza tra la tariffa intera e quella scontata è possibile scrivere:

$$C_u = T_i - T_s \quad (9)$$

Inoltre se, al contempo, si nota che anche il costo sostenuto per aver sovrastimato la domanda rappresenta un mancato guadagno esattamente pari alla tariffa scontata, in quanto non si è riusciti ad allocare parte delle risorse, rimaste poi invendute, avendole destinate e riservate per tutto l'orizzonte temporale delle prenotazioni ai clienti disposti a pagare la tariffa intera pur avendo avuto la possibilità di venderle a prezzo scontato, è possibile scrivere:

$$C_o = T_s \quad (10)$$

Riscrivendo la formula (8) utilizzando le relazioni (9) ed (10) appena individuate è possibile notare la convergenza dei risultati ottenuti attraverso l'utilizzo di questa tecnica e di quella basata sull'analisi decisionale.

Infatti:

$$\Pr \{ D_i < S \} \leq C_u / (C_u + C_o) \Leftrightarrow \Pr \{ D_i < S \} \leq (T_i - T_s) / T_i$$

4.7.5 Metodo risolutivo nel caso di più di due tariffe

La programmazione lineare è uno dei principali strumenti che permettono di supportare il processo di allocazione delle unità di servizio all'interno delle diverse classi tariffarie. Nel tempo si sono succeduti diversi modelli, che in alcuni casi hanno fornito valide risposte alle esigenze manageriali, mentre in altri non hanno potuto trovare una applicazione pratica.

La formulazione tipica del problema di programmazione lineare prevede come obiettivo la massimizzazione dei ricavi, considerando due tipologie di vincoli:

- **di disponibilità:** il numero delle unità di servizio riservate a ciascuna classe tariffaria non deve eccedere nella sua globalità, la capacità di offerta;
- **di domanda:** il numero delle unità riservate a ciascuna classe tariffaria non deve superare la domanda prevista.

Il problema, in termini di programmazione lineare semplice, può essere formulato come:

$$\max \quad \sum_i (r_i * x_i) \quad \text{con } i = 1, \dots, n;$$

$$\text{s.t.} \quad \sum_i x_i \leq C \quad \text{con } i = 1, \dots, n; \quad (\text{Vincolo 1})$$

$$x_i \leq d_i \quad \text{con } i = 1, \dots, n; \quad (\text{Vincolo 2})$$

$$x_i, d_i \geq 0 \quad \text{con } i = 1, \dots, n; \quad (\text{Vincolo 3})$$

Per comprendere la formulazione del modello si introducono i seguenti parametri:

- i pedice indicante la classe di prezzo o tariffaria;
- x_i variabile decisionale che rappresenta il numero di unità di servizio da riservare alla classe i -esima;
- d_i di domanda attesa relativa alla classe di prezzo i ;
- r_i ricavo unitario derivante dalla vendita di un'unità della classe i ;
- C capacità offerta espressa in unità di servizio.

La funzione obiettivo consiste nella massimizzazione del ricavo ottenuto vendendo ogni unità al più alto tra i prezzi praticabili.

Il primo vincolo stabilisce che il totale delle prenotazioni accettate non può superare la capacità massima di servizio mentre il secondo impone che le prenotazioni non possano eccedere la domanda. Infine l'ultimo vincolo stabilisce che sia la domanda che il numero di prenotazioni devono sempre essere positivi, infatti, non avrebbero alcun senso logico e matematico se fossero negativi.

Il modello quindi, a partire dalla conoscenza della capacità, della domanda e del ricavo unitario per ogni classe individua la migliore allocazione possibile della capacità di cui si dispone alle diverse classi tariffarie.

Il metodo di funzionamento del modello, in termini di variabili in ingresso ed in uscita, è illustrato nella figura seguente:

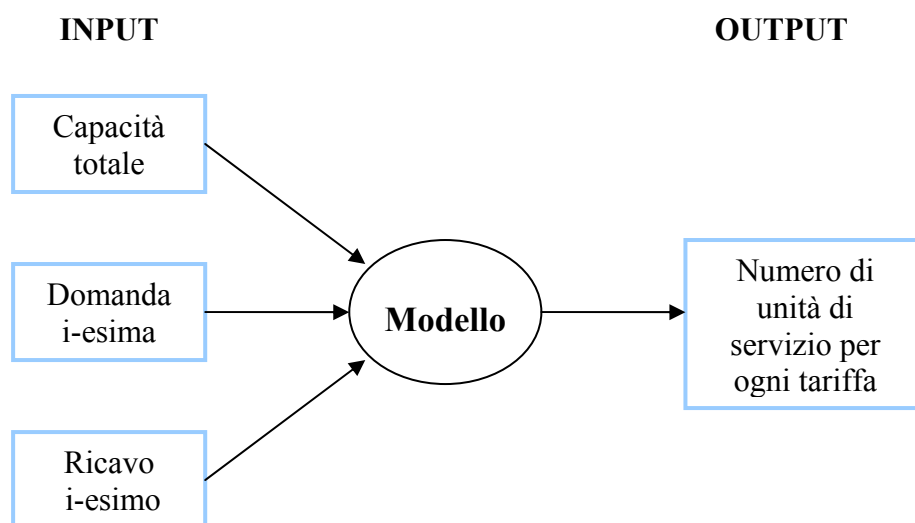


Figura 4.6: Inputs ed output del modello 1

Sebbene il problema sia di facile risoluzione, la sua applicazione risulta poco utile, poiché l'assunzione della domanda deterministica rende il risultato ottenuto poco aderente alla realtà. Il modello presuppone infatti la perfetta conoscenza della domanda totale espressa per ogni classe tariffaria e, nonostante il rigore insito nei modelli previsionali, una conoscenza perfetta è certamente impensabile. Inoltre non vengono tenuti in considerazione i fenomeni che modificano le previsioni fatte quali no-shows e cancellazioni.

Il modello bid – price rappresenta un'evoluzione del modello di programmazione lineare in quanto tenta di superare i limiti di quest'ultimo, responsabili della sua scarsa applicabilità e, per tali ragioni, sta ottenendo larghi consensi, in particolare presso le industrie turistiche.

Nello specifico nel modello vengono considerati una serie di fenomeni non facilmente prevedibili ma che, in ogni caso, modificano anche in maniera considerevole l'allocazione delle risorse quali, ad esempio, no-show, cancellazioni improvvise o show-up.

Il modello ha come obiettivo la definizione del numero delle unità di servizio da riservare a ciascuna classe di prezzo in modo da massimizzare i ricavi in un determinato istante temporale, tenendo conto, al contempo, dei vincoli connessi alla disponibilità di unità di offerta ed alle previsioni di domanda.

Il modello si articola in due fasi: la prima volta a determinare il limite massimo di prenotazioni da accettare per ogni classe tariffaria; la seconda, quale logica estensione della precedente, definisce l'iter decisionale di accettazione o di rifiuto di una richiesta di prenotazione.

Il modello di programmazione lineare è del tipo:

$$\max \sum_i (p_i r_i x_i) \quad \text{con } i = 1, \dots, n;$$

$$\text{s.t. } \sum_i p_i x_i \leq C + O \quad \text{con } i = 1, \dots, n; \quad (\text{Vincolo 1})$$

$$x_i \leq d_i \quad \text{con } i = 1, \dots, n; \quad (\text{Vincolo 2})$$

$$x_i, d_i \geq 0 \quad \text{con } i = 1, \dots, n; \quad (\text{Vincolo 3})$$

Per comprendere la formulazione del modello si introducono i seguenti parametri:

- i rappresenta la classe di prezzo o tariffaria;
- x_i è la variabile di scelta che rappresenta in numero di unità di servizio da riservare alla classe i -esima;
- d_i domanda attesa relativa alla classe di prezzo i ;
- r_i ricavo unitario derivante dalla vendita di un'unità della classe i ;

- p_i probabilità di show-up associata alla classe i ;
- C capacità offerta espressa in unità di servizio;
- O numero delle unità di offerta in overbooking.

Le variabili in ingresso e quelle in uscita, per le quali si implementa il modello, possono essere agevolmente schematizzate nelle figura seguente.

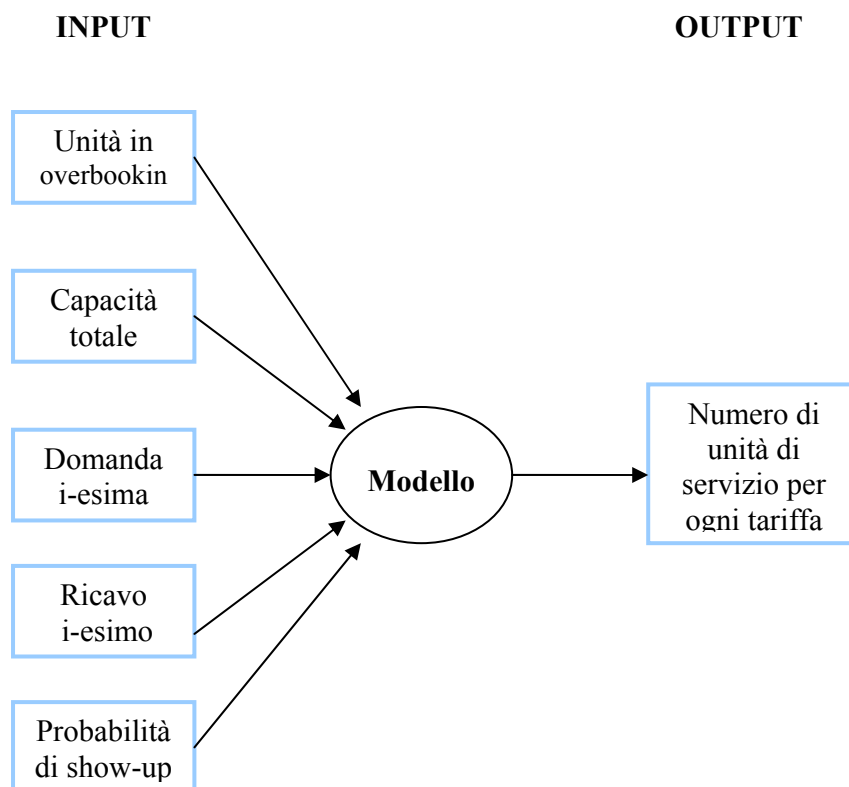


Figura 4.7: Inputs ed output del modello 2

Anche in questo caso il modello fornisce la soluzione ottimale per l'allocazione delle unità di servizio alle diverse classi tariffarie; al contempo si determinano i prezzi ombra⁷ associati ad ogni vincolo, ossia i valori incrementali della funzione obiettivo a fronte di un incremento del membro destro del vincolo stesso, dovuti ad esempio a no-show. I prezzi ombra rappresentano pertanto i ricavi incrementali che si otterrebbero avendo un'unità di offerta in più a disposizione. L'imprenditore è disposto ad accettare una prenotazione solo se il ricavo realizzabile con essa eccede quello a cui si rinuncia per la sua detrazione dalla disponibilità complessiva, noto in letteratura come *displaced revenue*. Il ricavo atteso è quindi il ricavo marginale connesso alla vendita dell'unità conseguente

⁷ In inglese Bid Price, da cui il nome del modello.

all'accettazione della richiesta di servizio mentre il ricavo perso per la sottrazione dell'unità è il costo atteso conseguente all'accettazione stessa.

Per individuare le soluzioni ottime si determina il ricavo supplementare che si otterrebbe avendo, per le future richieste, un'unità in più di servizio a disposizione. Nella formulazione del problema in p.l. tale elemento, rappresenta il prezzo ombra sul vincolo di capacità.

Si ipotizzi, a livello esemplificativo, una richiesta di prenotazione la cui probabilità di cancellazione e di no-show sia nulla. Occorre definire il ricavo realizzabile dall'accettazione r_a , che corrisponde al ricavo unitario associato alla specifica classe tariffaria ed il ricavo a cui si rinuncia r_d :

$$r_a = r_i$$

$$r_d = p_o$$

con p_o che è uguale al prezzo ombra nella soluzione del problema.

Se $r_i \geq p_o$, la richiesta di prenotazione deve essere accettata.

In termini di classi tariffarie ciò significa che tutte le classi pari o superiori al bid price sono aperte mentre quelle inferiori devono essere chiuse.

4.8 Concetto di overbooking

Le scelte riguardo all'overbooking risultano più complesse in quanto si deve inizialmente scegliere se prevedere questa opportunità e, in caso affermativo, stabilire il numero di unità da allocare in overbooking.

Nella prima decisione, scegliendo di praticare l'overbooking, si accetta in quel giorno un numero di prenotazioni superiore alla disponibilità fisica della struttura (aereo, albergo, sala cinematografica, nave, etc.), mettendosi così al riparo dagli eventuali mancati guadagni dovuti alle prenotazioni annullate ed ai *no-show*; questi ultimi sono i clienti che prenotano, senza poi presentarsi al check-in il giorno previsto. Di controparte, se tutti i prenotati si presentassero, si rischia di dover negare ad alcuni di essi la disponibilità.

Non praticando overbooking, il rischio di dover rifiutare un cliente prenotato non esiste, si potrebbero, però, perdere i profitti di coloro che pur avendo un posto riservato non si presentano, ovvero dei *no-show*. Nell'altra decisione, invece, se si scegliesse di vendere subito ed a prezzi scontati probabilmente non ci si ritroverebbe con delle risorse invendute, ma si rischierebbe di dover rifiutare una prenotazione a prezzo superiore (*rischio di rifiuto*), mentre, nel caso in cui si decidesse di aspettare a vendere, si potrebbe rischiare di ottenere delle risorse invendute (*rischio d'invenduto*). L'assunzione di una di queste politiche e del relativo rischio è definita dopo un'accurata analisi dei dati storici, delle previsioni future ed anche in base all'esperienza del management [22]. Il termine overbooking deriva dai termini inglese *over* (sopra) e *booking* (prenotazione) e può essere tradotto in italiano come "sovraprenotazione". La politica di overbooking fa esplicito riferimento alla gestione della capacità produttiva e si sostanzia nell'accettazione di un numero di prenotazioni superiore all'effettiva disponibilità di unità di servizio o nella vendita di un numero di biglietti superiore alla capacità effettiva al fine di tutelarsi da eventuali improvvise

mancate fruizioni del servizio stesso. L'overbooking, infatti, può essere applicato in riferimento a due diversi momenti decisionali e si distingue in:

- Overbooking sulle prenotazioni;
- Overbooking sulle vendite o *oversales* [18].

Fare overbooking sulle prenotazioni significa accettare un numero di prenotazioni superiore rispetto a quelle che l'azienda è in grado di fornire ed è effettuato con il fine di tutelarsi dalle prenotazioni che non si trasformano in acquisti, non sempre disdette con largo anticipo. Fare overbooking sulle vendite significa, invece, vendere contemporaneamente il servizio a più persone ed è effettuato con il fine di tutelarsi da eventuali no-show, ossia di mancate presentazioni al momento della fruizione del servizio, pur avendone precedentemente acquistato il diritto.

E' opportuno notare che applicando l'overbooking sulle prenotazioni si identificano dei valori dello stesso relativi ad un preciso istante temporale in quanto il concetto è dinamico ed i valori mutano da istante ad istante, così come le prenotazioni e le vendite. L'andamento dell'overbooking pertanto, essendo questo una percentuale della capacità prenotata, non è costante ma diminuisce nel tempo al passare della capacità da prenotata a venduta, come illustrato nella figura seguente:

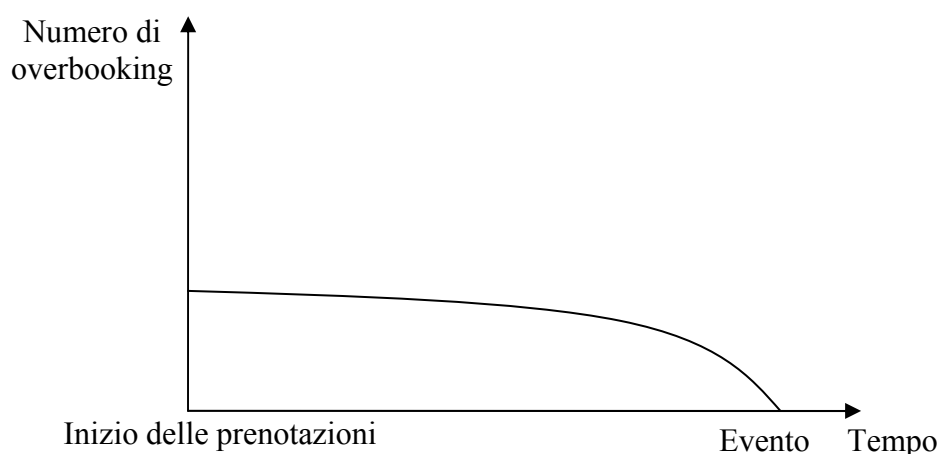


Figura 4.8: Andamento dell'overbooking sulle prenotazioni

Applicare una politica di overbooking significa sia applicare l'overbooking sulle prenotazioni o l'overbooking sulle vendite sia applicarli entrambi; la scelta dipende esclusivamente dalla tipologia di azienda che si sta analizzando.

Se la fruizione del servizio è preceduta da due fasi distinte dal punto di vista temporale, una in cui si effettua la prenotazione del servizio stesso ed una in cui si acquista il biglietto, si praticheranno entrambi i tipi di overbooking. Questo processo, tipico degli alberghi e delle sale cinematografiche, è illustrato graficamente nella figura seguente:

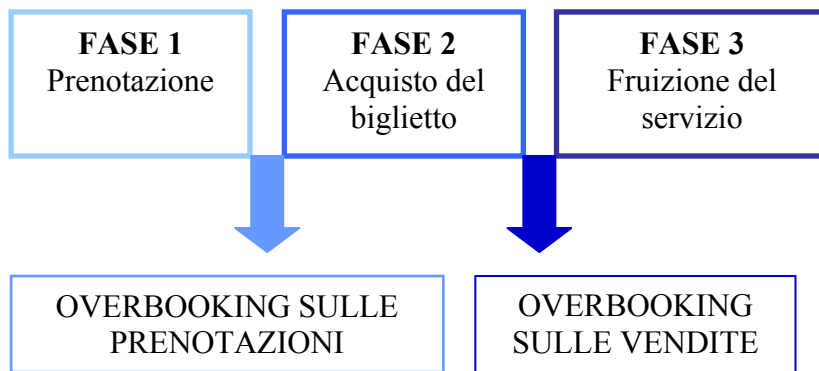


Figura 4.9: Processo in tre fasi

Se invece la fruizione del servizio è preceduta da un'unica fase in cui si prenota il servizio e al contempo si acquista il biglietto si praticherà soltanto un overbooking sulle vendite. Questo processo, tipico delle compagnie aeree, è rappresentato nella figura che segue:

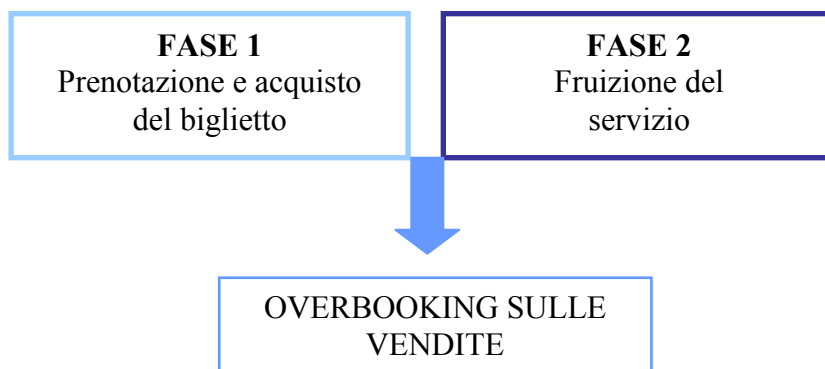


Figura 4.10: Processo in due fasi

Sia che si tratti di overbooking sulle prenotazioni che di overbooking sulle vendite l'obiettivo è il medesimo e consiste nel saturare in ogni istante la capacità produttiva e contenere da un lato il rischio di mancata vendita e dall'altro quello derivante da possibili disdette delle prenotazioni. L'andamento dell'overbooking, in relazione alla capacità, è il seguente:



Figura 4.11: Overbooking nel tempo

Vari sono i fattori da tenere in considerazione nella definizione del numero massimo di prenotazioni da accettare:

- possibilità di no-show o cancellazioni che originano una mancata vendita;
- costo opportunità della piena occupazione;
- valore connesso alla perdita di un cliente;
- reali possibilità di offrire sistemazioni alternative;
- costi delle sistemazioni stesse.

Un sistema efficiente deve essere in grado di bilanciare i fattori di cui sopra al fine di ottenere il migliori risultato possibile, sia in termini economico/reddituale che in termini di immagine.

L'azienda operando applicando l'overbooking indubbiamente consegue dei ricavi aggiuntivi in quanto riesce sempre a saturare la propria capacità ma, al contempo, potrebbe sostenere anche dei costi aggiuntivi, scomponibili in due aliquote: una finanziaria ed una non finanziaria.

Il costo finanziario che l'azienda sostiene è rappresentato dal cosiddetto "recupero del servizio" cioè dall'indennizzo che essa è disposta ad accordare al cliente fatto oggetto di questa pratica, strettamente dipendente dalla politica commerciale dell'azienda [23]. In altre parole il recupero del servizio rappresenta l'insieme delle azioni intraprese da un'azienda per gestire un insuccesso nell'erogazione di un servizio. Ad esempio è consuetudine riconoscere al cliente che si presenta al check-in e per il quale non è più disponibile l'unità di servizio, una sistemazione alternativa dotata di medesimi standard qualitativi o perfino di qualità superiore (*upgrading*); tutti i costi aggiuntivi di questa operazione, detti costi di riprotezione, sono completamente a carico dell'azienda [24].

Il costo non finanziario si associa alla caduta di immagine dell'azienda ed è originato dall'impossibilità di soddisfare la richiesta di un cliente malgrado la sua prenotazione: un cliente riprotetto presso altre strutture e gestito maldestramente è un cliente perso per il futuro, soprattutto qualora sia un cliente abituale. Questo gap inficia pesantemente lo standard di qualità.

Il management deve quindi gestire con attenzione l'overbooking che, se utilizzato opportunamente, può rappresentare una fonte di guadagno aggiuntiva ma che, se gestito in modo errato, potrebbe portare addirittura ulteriori costi all'azienda.

I costi aggiuntivi che l'overbooking porta all'azienda, in caso di previsioni errate, sono ipotizzati lineari per motivi pratici, benché come mostra la seguente figura il loro andamento non sia tale.

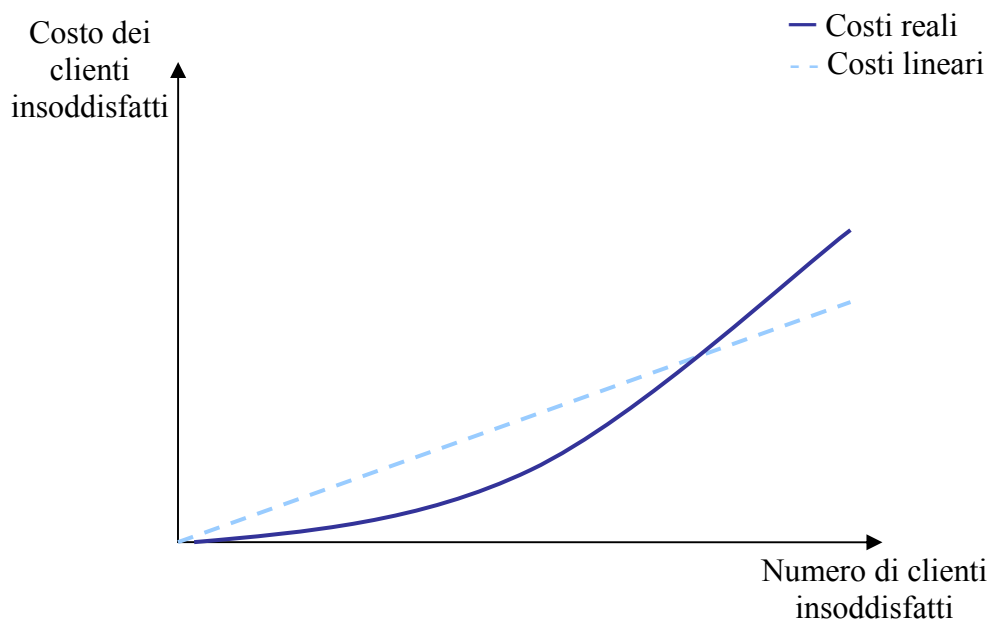


Figura 4.12: Costi reali connessi all'overbooking

Pertanto è possibile concludere che i ricavi aggiuntivi ottenibili attraverso l'utilizzo dell'overbooking sono crescenti in modo direttamente proporzionale alle risorse allocate fino al raggiungimento della saturazione, mentre dopo aver raggiunto la massima occupazione ed aver saturato la capacità di cui si dispone si trasformano in costi supplementari che, come detto, non sono linearmente decrescenti ma hanno un andamento decrescente più spinto, di tipo quadratico.

Quanto appena affermato è illustrato nella figura seguente:

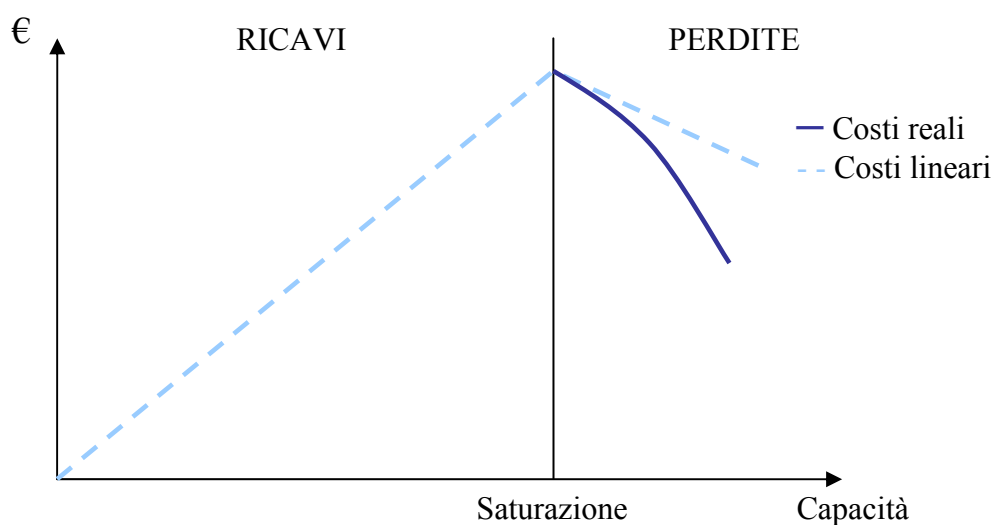


Figura 4.13: Andamento reale dei ricavi e dei costi

La gestione dell'overbooking si presenta quindi, alla luce delle considerazioni fatte finora, estremamente complessa.

Ad essa si collegano inoltre, come già detto, una serie di problematiche non di poco conto che portano il cliente al quale è stato negato il servizio, perché in overbooking, ad allontanarsi dall'azienda e a non usufruire più dei suoi servizi.

L'azienda che pratica overbooking e che non lo gestisce opportunamente, ritrovandosi con più clienti di quanti ne può effettivamente servire, tradisce le aspettative di questi ultimi e mina inesorabilmente il rapporto di fiducia che si è venuto a creare nel tempo. Per questa ragione ha un peso rilevante il tema del recupero del servizio, atto proprio a tentare di recuperare il rapporto di fiducia perso e l'immagine aziendale decaduta [25]. L'obiettivo delle aziende che praticano overbooking resta comunque quello di massimizzare i ricavi ma non a qualunque costo in quanto consapevoli dei maggiori danni che possono derivare da situazioni di overbooking.

In base alle considerazioni fatte finora una corretta gestione dell'overbooking dovrebbe rispondere ai seguenti principi [26]:

- essere su base volontaria. Si ricorda a questo proposito l'esperienza nel settore aereo del “*voluntary bump*”. In pratica, anziché procedere alla determinazione unilaterale del cliente cui negare il servizio prenotato, l'impresa cerca di trovare alcuni clienti disposti ad accettare spontaneamente il cambiamento delle condizioni di erogazione del servizio in cambio di una qualche forma di compensazione. E' infatti evidente che il timing nella fruizione di un determinato servizio può essere cruciale per alcuni individui mentre per altri risulta meno importante o persino irrilevante;

- fornire un'adeguata e tempestiva informazione in modo da permettere eventualmente ad alcuni clienti di esplorare soluzioni alternative e gestire poi, per conto del cliente, i cambiamenti nelle modalità di fruizione del servizio;
- eventualmente ricordare al cliente e ai dipendenti che un beneficio dell'overbooking è dato dal fatto che in questo modo si riesce ad incrementare il numero dei clienti serviti e potere offrire i servizi a prezzi più bassi;
- il servizio di recupero dovrebbe evitare, per quanto possibile, di portare il cliente a fruire dei servizi di un concorrente dal momento che quest'ultimo potrebbe approfittare della circostanza per agganciare il nuovo cliente.

4.9 Disciplina normativa

La rilevanza economica e giuridica dell'overbooking ha posto la necessità di cercare delle soluzioni normative adeguate.

Le leggi a tutela dei consumatori che subiscono i disagi dovuti all'utilizzo di questa pratica da parte dell'azienda sono numerose: la più importante consiste nel rimborso di denaro da parte dell'azienda ai consumatori per il servizio acquistato di cui non si è potuto usufruire o, in alternativa, nell'offerta, a carico dell'azienda, di un servizio alternativo di qualità pari o superiore a quello acquistato atto a soddisfare il medesimo bisogno.

Il fenomeno, però, è diffuso soprattutto nel settore aereo, e proprio per regolamentare questo settore così complesso dal 17 febbraio 2005 vige un nuovo ed analitico regolamento: il Regolamento CE n. 261 dell'11 febbraio 2004.

Il Regolamento disciplina l'utilizzo dell'overbooking, ovvero contiene l'insieme delle regole poste a protezione del consumatore nel caso di diniego di imbarco su un aereo per overbooking (eccesso di prenotazioni rispetto ai posti disponibili sull'aereo), e per altri casi non riguardanti questa trattazione, quali cancellazioni o ritardi prolungati dei voli. Il Regolamento, quale atto normativo comunitario di immediata applicazione negli Stati membri, non necessita di un atto normativo nazionale di recepimento, rafforzando in modo sostanziale la tutela del viaggiatore a cui venga negato l'imbarco su un aereo a causa di un eccesso di prenotazioni rispetto ai posti disponibili, sia in termini di risarcimento pecuniario che in termini di assistenza da parte del "vettore", cioè della compagnia aerea che organizza il volo.

Il nuovo Regolamento si applica sia ai voli di linea che ai voli charter (a differenza dei precedenti regolamenti), da soli o inseriti in un pacchetto turistico "tutto compreso" (quelli che comprendono servizi di alloggio, trasporto ed altri servizi turistici non accessori ai primi due).

I passeggeri che hanno diritto alla tutela sono quelli che dispongono di una prenotazione confermata sul volo in cui viene negato l'imbarco e che si sono presentati all'accettazione all'ora indicata per iscritto o, in mancanza di questa, almeno 45 minuti prima dell'ora di partenza (articolo 3).

La novità del Regolamento è rappresentata dall'obbligo per la compagnia aerea, prima di negare l'imbarco al viaggiatore per eccesso di prenotazioni, di interpellare i passeggeri per trovare eventuali volontari disposti a

rinunciare al volo in cambio di un risarcimento e a negare l'imbarco a passeggeri non consenzienti solo qualora il numero dei volontari non sia sufficiente a consentire l'imbarco a tutti quelli muniti di prenotazione (art. 4). Inoltre, è stata aumentata, rispetto alle precedenti norme, l'entità della compensazione pecuniaria da riconoscere nel caso di overbooking al passeggero⁸ che continua ad usufruire, in aggiunta ad essa, del diritto alla scelta tra il rimborso del biglietto ed un volo alternativo per la stessa destinazione o un'altra ad essa vicina, oltre ai pasti ed alla sistemazione alberghiera (ed al trasporto verso e da essa) per il tempo necessario a trovare una soluzione al suo mancato trasferimento alla destinazione desiderata.

Se il volo sostitutivo è in una classe superiore, il vettore non ha diritto alla differenza di prezzo col biglietto del volo sostituito, se è inferiore il passeggero ha invece diritto al rimborso che varia dal 30% al 75% del prezzo del biglietto a seconda della tratta dell'aereo.

Nel caso in cui la compagnia aerea o il tour operator che hanno prenotato il volo lo cancellino per loro responsabilità (art. 5), i passeggeri hanno diritto ai risarcimenti sopra indicati a meno che:

- siano stati informati della cancellazione del volo almeno due settimane prima;
- siano informati con congruo anticipo e riprotetti (cioè riprenotati) con una partenza fissata ad un orario molto vicino (una o due ore) a quello originario.

L'onere della prova sul se e quando il passeggero è stato avvertito della cancellazione del volo spetta al vettore aereo. Se la cancellazione del volo avviene per "circostanze eccezionali" (cioè per cause di forza maggiore) non è dovuta la compensazione pecuniaria (Fonte: www.ecc-netitalia.it/viaggiturismo).

I diritti dei passeggeri stabiliti dal Regolamento n. 261 del 2004 sono irrinunciabili, anche nel caso di clausole derogatorie presenti nel contratto di trasporto le quali lasciano, in ogni caso, impregiudicati gli eventuali risarcimenti supplementari.

Il Codice Civile stabilisce con l'articolo 2951 che i diritti che derivano dal contratto per il trasporto di persone entrano in prescrizione dopo un anno di tempo; ciò vuol dire che presentando il contratto di viaggio in originale, i clienti hanno 12 mesi per richiedere un risarcimento dei danni ulteriori.

Infatti, oltre a quanto previsto dal Regolamento CE il passeggero può chiedere alla compagnia aerea il risarcimento dei danni subiti in conseguenza del mancato imbarco sul volo da lui prenotato: ad esempio danno patrimoniale, nei due profili del danno emergente (spese da affrontarsi in conseguenza del mancato imbarco sul volo prenotato: taxi, pernottamento in albergo...) e del lucro cessante (mancato guadagno per la non conclusione di un affare, perdita di ore di lavoro, mancato godimento di alcuni dei giorni di vacanza), nonché della perdita di una chance (eventuale occasione di lavoro sfumata), oltre al danno biologico da stress per i disagi subiti.

In ottemperanza al Regolamento CE n.261 del 2004 il vettore aereo ha l'obbligo di informare per iscritto i passeggeri in merito ai loro diritti in tutti i casi di negato imbarco, cancellazione o ritardo del volo di almeno due ore

⁸ Il risarcimento per negato imbarco ammonta, con il Reg. 261 a: 250 euro per i voli inferiori o pari a 1500 Km (di rotta ortodromica), 400 euro per i voli compresi tra 1500 e 3500 Km e 600 euro per tutti gli altri voli.

ed inoltre, ogni Stato membro dell'Unione Europea deve designare un "organismo" (preferibilmente pubblico o anche privato) responsabile dell'applicazione del suddetto regolamento per i voli da e per gli aeroporti situati nel suo territorio e presso cui i passeggeri possano presentare reclami per le presunte violazioni dei diritti loro riconosciuti da questo atto normativo. Detto organismo adotta tutte le misure necessarie, comprese quelle sanzionatorie di tipo amministrativo stabilite dagli Stati membri, per garantire che siano rispettati i diritti dei passeggeri-consumatori.

Il consumatore, destinatario della direttiva, nei contratti di trasporto può essere definito come: "qualsiasi persona fisica che agisce per fini che non rientrano nel quadro della sua attività professionale" (Fonte: Direttiva CE del 5 aprile 1993, n.13, art.2.); pertanto egli è una persona che non persegue scopi imprenditoriali ma che svolge un'attività diretta al soddisfacimento di esigenze personali, proprie o della propria famiglia, intendendo per attività anche il mero contatto sociale.

Spetta all'utente, in caso si verificassero dei problemi, dimostrare l'esistenza delle proprie caratteristiche di consumatore, cioè di soggetto che stipula, direttamente o tramite rappresentante, il contratto di trasporto per scopi estranei alla sua attività professionale o imprenditoriale.

Il vettore è invece semplicemente considerato come un imprenditore secondo quanto previsto dall'art. 1469 bis del Codice Civile.

4.10 Metodi risolutivi

I diversi metodi risolutivi esistenti hanno tutti il medesimo obiettivo che consiste nell'individuare il miglior numero di unità da allocare in overbooking.

Il compito consiste nel trovare un giusto compromesso tra i costi che l'azienda sostiene per non aver venduto tutte le risorse di cui dispone e i costi di riprotezione che l'azienda sostiene nel momento in cui alloca più unità di cui dispone e tutti i clienti si presentano per usufruire del servizio. Il risultato di quest'analisi è rappresentato dal livello ottimo dell'overbooking.

4.10.1 Analisi sui valori medi

L'analisi sui valori medi è una tecnica molto semplice, intuitiva e di facile implementazione.

Il livello ottimo dell'overbooking viene determinato attraverso una semplice equazione:

$$BL = C / p$$

dove:

- BL, acronimo di *Booking Limit*, indica il livello ottimo dell'overbooking e rappresenta l'incognita di questo tipo di problemi;
- C rappresenta la capacità totale di cui l'azienda dispone;
- p rappresenta la probabilità che un cliente che ha fatto una prenotazione si presenti per usufruire del servizio che ha precedentemente prenotato.

La tecnica, benché molto semplice ha il difetto di non considerare i costi di riprotezione per cui risulta adatta solo ai contesti nei quali questi ultimi sono di basso valore.

4.10.2 Analisi tramite foglio di calcolo

L'analisi tramite foglio di calcolo è una tecnica immediata e può essere implementata sia attraverso l'ausilio di una semplice tabella o, nel caso in cui bisogna analizzare un elevato numero di variabili, attraverso un foglio di lavoro Excel.

Questa tecnica supera gli svantaggi dell'analisi sui valori medi infatti considera i costi connessi all'overbooking, ma di quest'ultima mantiene il principale vantaggio cioè la semplicità.

I costi considerati sono di due tipi:

- *Spoilage Cost*: costo che l'azienda sostiene quando non sfrutta una risorsa, che rappresenta un costo opportunità;
- *Walking Cost*: costo che l'azienda sostiene quando ad un cliente con prenotazione non è garantito il servizio.

Una volta che il management ha definito i valori associati a questi due costi si costruisce una tabella nella quale si inseriscono i no-show con le relative probabilità di accadimento e si inseriscono i valori dello spoilage cost (C_u) e del walking cost (C_o) :

Numero di prenotazioni in overbooking									
No-show	Probabilità	0	1	2	3	4	5	6	7
0	0.05	0	Co	2*Co	3*Co	4*Co	5*Co	6*Co	7*Co
1	0.10	Cu	0	Co	2*Co	3*Co	4*Co	5*Co	6*Co
2	0.20	2*Cu	Cu	0	Co	2*Co	3*Co	4*Co	5*Co
3	0.25	3*Cu	2*Cu	Cu	0	Co	2*Co	3*Co	4*Co
4	0.20	4*Cu	3*Cu	2*Cu	Cu	0	Co	2*Co	3*Co
5	0.10	5*Cu	4*Cu	3*Cu	2*Cu	Cu	0	Co	2*Co
6	0.05	6*Cu	5*Cu	4*Cu	3*Cu	2*Cu	Cu	0	Co
7	0.05	7*Cu	6*Cu	5*Cu	4*Cu	3*Cu	2*Cu	Cu	0
Costo totale									

Tabella 4.1: Individuazione del valore ottimo di overbooking

Nell'ultima riga si calcola poi, per ogni colonna, il costo atteso totale, ottenuto moltiplicando ogni costo della colonna per la relativa probabilità di accadimento e facendone la somma. Tra i vari valori ottenuti si sceglie quello a cui corrisponde il costo atteso minimo ed infine si associa alla colonna individuata il numero ottimo di prenotazioni in overbooking da accettare al fine di massimizzare il ricavo [18].

I vantaggi di questa tecnica sono diversi: risulta di semplice implementazione, i dati di cui necessita sono di semplice reperimento e la forma tabellare in cui questi ultimi sono raccolti facilita sia l'ordine che il loro corretto inserimento. Inoltre il valore ottimo è individuato attraverso un opportuno bilanciamento dei costi opportunità e dei costi da sostenere, cioè in modo coerente con l'obiettivo dell'overbooking.

Lo svantaggio di questa tecnica risiede unicamente nella sua forma tabellare, che, se da un lato, come detto, rappresenta un pregio in quanto permette un'efficace organizzazione dei dati di cui si dispone, dall'altro rappresenta un difetto in quanto non permette di comprendere a fondo il meccanismo che sta alla base della soluzione che viene presentata semplicemente attraverso una serie di calcoli matematici. In conclusione è possibile affermare che la tecnica di analisi attraverso il foglio di calcolo risulta indubbiamente migliore della tecnica di analisi sui valori medi in quanto, a fronte di un piccolissimo incremento della complessità di implementazione, conduce ad una soluzione certamente più precisa ed accurata che viene individuata attraverso un'analisi incentrata sui costi, elementi importanti per la determinazione del valore ottimo di overbooking.

4.10.3 Analisi tramite albero delle decisioni

L'albero delle decisioni consente la visualizzazione di un'analisi decisionale cioè dell'iter logico che sta dietro una certa decisione, l'analisi delle possibili alternative e delle conseguenze di ognuna di queste.

L'albero decisionale è rappresentato da quadrati o rettangoli che rappresentano le decisioni, da frecce che rappresentano le alternative più importanti e da cerchi che rappresentano i nodi probabilistici i cui risultati sono incerti.

Anche la scelta dell'accettazione della prenotazione di un cliente in overbooking può essere rappresentata tramite un albero delle decisioni come il seguente:

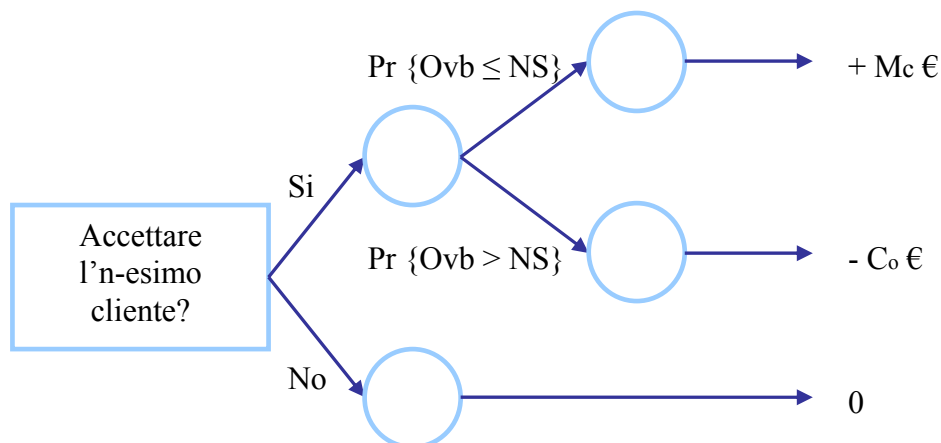


Figura 4.14: Albero decisionale per un problema di overbooking

Dall'albero decisionale è possibile comprendere che è conveniente accettare il cliente n-esimo solo se la probabilità di overbooking è minore o pari alla probabilità di no-show: in questo caso si ottiene un ricavo aggiuntivo pari ad Mc . Nel caso in cui, invece, si decide di accettare il cliente n-esimo nonostante la probabilità di overbooking sia maggiore di quella di no-show si rischia una perdita, quantificabile nel costo Co che l'azienda deve sostenere per rimborsare il cliente non servito.

Dal punto di vista matematico, con la stessa logica probabilistica, è possibile affermare che è opportuno e conveniente accettare il cliente n-esimo se si verifica che:

$$Mc * Pr \{ Ovb \leq NS \} \geq Co * Pr \{ Ovb > NS \} \quad (3.1)$$

da cui ricordando che la probabilità totale ha valore unitario:

$$Mc * Pr \{ Ovb \leq NS \} \geq Co * [1 - Pr \{ Ovb \leq NS \}] \quad (3.2)$$

ed infine, isolando la probabilità ad un membro:

$$Pr \{ Ovb \leq NS \} \geq Co / (Mc + Co) \quad (3.3)$$

Il valore di overbooking ottimo viene poi individuato scegliendo il più grande valore di n che soddisfa la disuguaglianza (3.3).

Questa tecnica di analisi, a differenza dell'analisi sui valori medi e dell'analisi tramite foglio di calcolo, presenta diverse difficoltà insite sia nella stima delle varie probabilità che nell'implementazione stessa. Diversamente dalle suddette fornisce però, oltre alla soluzione ottima, una serie di informazioni aggiuntive ed inoltre vengono evidenziate le varie decisioni prese che hanno poi condotto alla scelta finale rendendo quest'ultima più chiara, logica e facilmente comprensibile.

4.10.4 Analisi marginale

L'analisi marginale è una tecnica molto generica e versatile in quanto è adatta a diversi contesti applicativi.

Nel caso specifico di determinazione del livello di overbooking con questa tecnica si continuano ad accettare prenotazioni fino a quando il margine atteso sull'ultima prenotazione accettata è superiore o pari alla perdita attesa sempre sull'ultima prenotazione.

In modo compatto indicando con:

- C_u il costo che l'azienda sostiene nel sottostimare il numero di posti in overbooking;
- C_o il costo che l'azienda sostiene nel sovrastimare il numero di posti in overbooking;

è possibile scrivere:

$$C_u * Pr \{ Ovb \leq NS \} \geq C_o * Pr \{ Ovb > NS \} \quad (3.4)$$

ricordando che la probabilità ha spazio campione unitario:

$$C_u * Pr \{ Ovb \leq NS \} \geq C_o * [1 - Pr \{ Ovb \leq NS \}] \quad (3.5)$$

ed infine, isolando i costi:

$$Pr \{ Ovb \leq NS \} \geq C_o / (C_u + C_o) \quad (3.6)$$

E' possibile notare la convergenza tra il risultato ottenuto con l'analisi marginale (3.6) e quello ottenuto con l'analisi decisionale (3.3). Infatti:

$$Pr \{ Ovb \leq NS \} \geq C_o / (M_c + C_o) \Leftrightarrow Pr \{ Ovb \leq NS \} \geq C_o / (C_u + C_o)$$

4.10.5 Modello del ricavo atteso

Il modello del ricavo atteso inserisce tra le variabili il costo associato all'overbooking. Quest'ultimo, come già precedentemente accennato, si compone sia di elementi diretti, quali l'onere necessario a trovare una sistemazione alternativa, che indiretti, associabili a perdite d'immagine e rinuncia a future possibilità di reddito.

Il ricavo atteso per ogni numero di prenotazioni accettato può essere trovato, in modo semplice, attraverso la soluzione della seguente espressione:

$$\text{Ricavo Atteso} = \sum_i p(i) \{ R(i) + D * (n - i) \}$$

dove:

- i rappresenta il numero di clienti con una prenotazione e che si presume effettivamente arriveranno;
- $p(i)$ indica la probabilità associata ad ogni i con un totale di n prenotazioni accettate;
- $R(i)$ rappresenta il ricavo corrispondente ad un numero di clienti serviti pari ad i ;
- D indica il deposito quota non rimborsabile in caso di no-show o cancellazione;
- $(n - i)$ rappresenta il numero di no-show e cancellazioni.

La soluzione di questo modello fornisce il valore di n che permette di massimizzare il ricavo atteso. Questo rappresenterà quindi il numero di prenotazioni da accettare.

4.10.6 Modello di Karaesmen e van Ryzin

L'overbooking è uno dei principali problemi cui si tenta di dare una soluzione applicando le tecniche fornite dallo yield management. La maggior parte degli autori che hanno affrontato questa problematica hanno considerato l'esistenza di un solo tipo di bene. Nella realtà invece la maggior parte delle aziende tende a soddisfare le svariate esigenze della clientela attraverso diverse categorie di beni che si possono considerare sostanzialmente sostitutivi.⁹

I modelli analizzati fino a questo momento non consideravano queste diverse possibilità di offerta ma si limitavano a trattare in maniera univoca la complessiva disponibilità di servizio senza prendere in considerazione le singole peculiarità. Il sistema di offerta nella realtà invece è complesso e caratterizzato da una pluralità di servizi alternativi e, per certi versi, una cultura orientata allo yield lo rende ancora più complesso.

La principale innovazione presente nell'approccio di Karaesmen e van Ryzin, è appunto quella di considerare la differenziazione dell'offerta di un'impresa e di applicare a questa una politica di overbooking congiunto. Si superano, in un qualche modo, le trattazioni indipendenti per ogni categoria di servizio.

⁹ Le varie tipologie di offerta sono presenti in molti settori: le linee aeree dispongono di posti in prima classe, in economy ed in business class, le strutture ricettive offrono camere suites, junior suites e standard ed infine le aziende che noleggiavano auto dispongono di diversi modelli.

Il modello è valido ed applicabile anche in contesti di piccole dimensioni, tipici della realtà italiana. Poiché in tal caso, in conseguenza delle modeste dimensioni di ogni impresa, sarebbe arduo trovare politiche efficienti ed economicamente convenienti da implementare in ogni singola struttura si possono considerare le singole strutture come parte di un unico sistema di offerta complesso. Le unità di servizio esistenti all'interno di ogni azienda si suppone costituiscano l'insieme di beni disponibili in una realtà di più grandi dimensioni. Questa considerazione permette di superare parte dei problemi conseguenti alle piccole dimensioni e di implementare ugualmente una politica di overbooking.

Si supponga di introdurre il modello in un sistema di offerta composto da m diverse classi di prenotazioni ed n diverse tipologie di servizi. L'impostazione originale proposta da Karaesmen e van Ryzin prevede una soluzione composta da due stadi [27].

Nel primo stadio definito *reservation period* o periodo delle prenotazioni, devono essere definite il numero delle prenotazioni u da accettare per ogni classe i . La variabile u_i può variare tra 1, che corrisponde alla classe più costosa, ed m individuante la classe più economica.

In questo primo stadio il problema può essere formulato nei termini seguenti:

$$\max G(u)$$

dove $G(u)$ viene definito come il valore corrispondente al ricavo atteso ed è funzione dei limiti di prenotazioni accettate $u_1+u_2+\dots+u_n$.

Il secondo stadio denominato *service period* o periodo di fruizione del servizio contraddistingue il momento di arrivo effettivo dei clienti. A questo punto è necessario tenere in considerazione eventuali no-show o cancellazioni e si confermano le prenotazioni associabili a clienti che effettivamente sono presenti. Questa variabile, relativamente alla classe i viene indicata con z_i . Il suo valore dipende, in primo luogo, dal numero di prenotazioni accettate e pertanto $z_i = z_i(u_i)$. La funzione ha evidentemente natura stocastica. Inoltre, come nella maggior parte dei modelli di overbooking si assume che la parte di prenotazioni che non giungono ad effettiva fruizione del servizio siano indipendenti ed abbiano una distribuzione uniforme. Da ciò consegue che il numero dei sopravvissuti per ogni classe $z_i(u_i)$ sia binomiale $((1-p_i), u_i)$, dove p_i rappresenta appunto la probabilità associata ai no-show ed alle cancellazioni. Una volta conosciuto $z(u)$, i clienti devono essere allocati ad una delle risorse disponibili (camere, posti a sedere, etc.) di tipo j , con j da varia da 1 ad n . Nuovamente 1 sarà il tipo servizio preferito mentre $n-1$ rappresenta quello meno preferito; alla classe n vengono assegnati i clienti che non si riescono a servire a causa dell'overbooking. Questo tipo di bene è quindi "virtuale" visto che i clienti allocati in n non verranno serviti direttamente dall'azienda. Il problema di ottimizzazione relativo al secondo stadio può essere posto nei termini seguenti:

$$\max \sum_i \sum_j a_{ij} y_{ij} \quad i=1,\dots, m \quad j=1,\dots, n$$

$$\text{s.t. } \sum_j y_{ij} = z_i \quad i = 1, \dots, m \quad (\text{vincolo 1})$$

$$\sum_i y_{ij} = c_j \quad j = 1, \dots, n \quad (\text{vincolo 2})$$

$$y_{ij} \geq 0 \quad i = 1, \dots, m \quad j = 1, \dots, n \quad (\text{vincolo 3})$$

avendo indicato con:

- a_{ij} il ricavo (costo) netto associabile all'assegnazione di un cliente prenotato in classe i ad un servizio appartenente alla classe di tipo j ;
- c_j il numero di risorse disponibili per ogni tipo j (cn individua il livello massimo di overbooking che si è disposti ad accettare);
- y_{ij} il numero di clienti prenotati in classe i assegnati ad una risorsa di tipo j ;
- z_i il numero di clienti in possesso di prenotazione in classe i che effettivamente si presentano.

Il primo vincolo impone semplicemente che ogni cliente effettivamente presentatosi per fruire del servizio sia anche servito.

Il secondo vincolo stabilisce che il numero di clienti assegnati ad ogni classe di servizio non possa essere maggiore della effettiva disponibilità.

Il terzo vincolo chiarisce semplicemente che il numero di clienti assegnati ad ogni classe di servizio non può mai essere negativo.

Occorre inoltre evidenziare che il parametro a_{ij} prende in considerazione tutti i costi ed i guadagni sia diretti che indiretti derivanti dal *down-grading* (cioè dall'assegnare un cliente prenotato per una classe superiore ad una classe inferiore) e dall'*up-grading* (cioè dall'assegnare un cliente prenotato per una classe inferiore ad una classe superiore). In questo parametro vengono pertanto inclusi anche la perdita o il guadagno di benevolenza (c.d. *goodwill*) da parte del cliente e la possibile perdita di un business futuro.

Il modello permette quindi una struttura di sostituzione generalizzata nella quale ogni tipologia di unità di servizio può essere utilizzata per provvedere a soddisfare la richiesta di acquisto relativa ad ogni classe di prenotazione. Ad ogni sostituzione posta in essere si ricollega un aumento o una diminuzione di redditività che viene adeguatamente introdotta nella ricerca della soluzione ottimale.

Nel modello si assume che ogni azienda, nel momento in cui decide come allocare le diverse risorse, conosca esattamente il numero di prenotazioni sopravvissute in tutte le classi. È evidente che questa può essere considerata solo un'approssimazione del mondo reale. L'assunzione è ritenuta comunque accettabile se si vuole rendere possibile uno sviluppo non eccessivamente problematico.

Il modello presenta alcuni limiti che possono essere posti nei termini seguenti:

- l'uso della distribuzione di Poisson come approssimazione alla distribuzione binomiale;
- l'uso di un algoritmo di tipo stocastic gradient;
- la mancanza di un metodo analitico per trovare il ricavo atteso e quindi per trovare la soluzione ottimale.

Gli autori sostengono che la distribuzione binomiale non si adatti ai metodi di ottimizzazione continua e scelgono la sua approssimazione con la distribuzione di Poisson. Questo implica l'eventualità di inserire soluzioni non realistiche. Il numero di prenotazioni sopravvissute $z_i(u_i)$ può essere maggiore del livello massimo di prenotazioni accettabili u_i . Parimenti le disdette ed i no-show $u_i - z_i(u_i)$, potrebbero assumere valori negativi. L'incoerenza è più facile a dimostrarsi quando il numero di unità di servizio disponibili non è molto elevato in quanto la ristrettezza del dominio rende meno accettabile l'approssimazione della distribuzione binomiale con quella di Poisson.

L'algoritmo di tipo *stochastic gradient* è un modo molto complesso per risolvere il problema. Risultati altrettanto buoni possono essere raggiunti in modo più semplice.

Il problema maggiore del modello presentato è la mancanza di un metodo analitico per calcolare il ricavo atteso e quindi trovare il risultato, mancanza solo parzialmente risolta ricorrendo alla simulazione. La ricerca di una soluzione migliore richiede l'introduzione di due concetti fondamentali: la proprietà di Monge e gli algoritmi di tipo greedy.

La proprietà di Monge fu scoperta dal matematico francese Gaspard Monge (1746 -1818).

Una matrice $C = (C_{ij})$, di dimensioni $m \times n$, soddisfa la proprietà di Monge se:

$$c_{ij} + c_{rs} = c_{is} + c_{rj} \quad \forall \quad 1 \leq i \leq r \leq m; 1 \leq j \leq s \leq n$$

Il modello scritto dallo studioso può essere reinterpretato come un problema standard di trasporto.

Nella figura sotto riportata si pone come obiettivo il trasporto nei punti a e b di due uguali quantità situate in A e B.

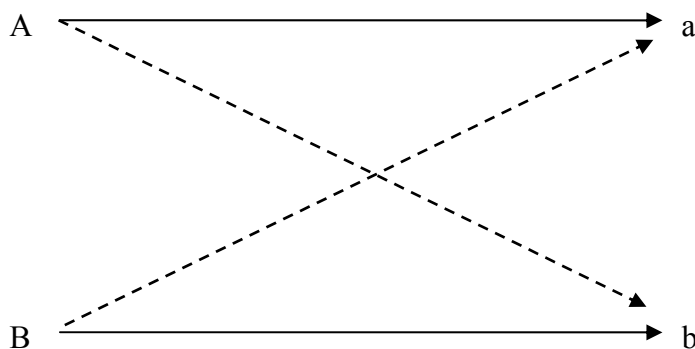


Figura 4.15: La proprietà di Monge

In una situazione di questo tipo Monge afferma che, se l'obiettivo è minimizzare i costi totali e se gli stessi sono direttamente proporzionali alla via percorsa, le linee aventi come origine A e B non devono mai intersecarsi.

Negli anni sessanta Alan J. Hoffman [28] [29] ha ripreso gli studi di Monge coniando il termine “proprietà di Monge” ed ha inoltre dimostrato che un problema di trasporto semplice la cui matrice dei costi soddisfa la suddetta proprietà può essere risolto usando un algoritmo di tipo greedy.

Un algoritmo di tipo greedy dal punto di vista matematico è un algoritmo costruttivo che individua iterativamente la soluzione definendo, ad ogni passo, l'elemento che, per quello specifico passo risulta più conveniente, senza fare alcun tipo di considerazione sulla composizione complessiva della soluzione. Con un algoritmo di tipo greedy, che letteralmente significa “affamato, avido”, può capitare pertanto che le ultime iterazioni risultino inefficienti dal momento che le possibilità di scelta sono molto ridotte [21].

4.11 Esempio applicativo: il caso di un ente di formazione

4.11.1 L'ente di formazione Ares e l'offerta formativa

Le iniziative di **formazione professionale** costituiscono un servizio di interesse pubblico inteso ad assicurare un sistema di interventi finalizzati alla diffusione di conoscenze teoriche e pratiche necessarie per svolgere ruoli lavorativi particolarmente richiesti dal mercato, rivolgendosi a diverse categorie: soggetti i quali hanno assolto l'obbligo scolastico ma non hanno mai svolto attività di lavoro, soggetti che lavorano, coinvolti in processi di specializzazione, riqualificazione e riconversione o, ancora, soggetti da lungo tempo senza lavoro in cerca di qualifica professionale.

In tale ambito opera l'associazione ARES fornendo due tipologie di servizi:

- il primo basato sui capitolati a seguito di bandi di gara indetti da enti europei, statali o regionali, e finalizzato allo sviluppo occupazionale e imprenditoriale di aree geografiche depresse;
- il secondo è costituito da servizi di formazione a commessa, indirizzati essenzialmente ad imprese interessate alla formazione, qualificazione o riqualificazione del proprio personale.

In entrambi i casi, i servizi offerti da ARES hanno come obiettivo lo sviluppo di conoscenze e capacità: le conoscenze riguardano il trasferimento di nozioni, procedure e principi generali specifici di una figura professionale, mentre lo sviluppo di capacità è finalizzato all'acquisizione di atteggiamenti e capacità legati alla risoluzione dei problemi lavorativi o di relazione interpersonale.

L'ARES è una associazione nata circa dieci anni fa con l'obiettivo sia di formare professionalmente che di offrire consulenza sia a giovani che a lavoratori sviluppando per essi programmi di intervento socio-formativi tali da renderli adeguati alle richieste del mondo del lavoro; pertanto esegue corsi di formazione per i giovani e riqualifica e riconverte i lavoratori in momentanea difficoltà.

Oltre a questo obiettivo principale l'Ares offre anche altri servizi tra cui di rilevante importanza è l'affiancamento e il supporto alle aziende, agli enti locali, alle associazioni ed agli organismi, che decidono di intraprendere nuove strade per lo sviluppo e l'innovazione.

Il fine ultimo dell'ARES, in ogni caso, è quello di rispondere, attraverso una consulenza specialistica qualificata, alle esigenze del qualsiasi cliente si presenti (azienda o persona, studente o lavoratore che sia) utilizzando moderne tecniche di analisi, di programmazione, di addestramento e di assistenza, formulando infine soluzioni integrate e modulari ad hoc per lo specifico caso.

L'offerta di servizi si concentra soprattutto in una vasta scelta di corsi professionali (da effettuare sia in aula con docenti specializzati, professionisti ed esperti dell'argomento, sia "sul campo" attraverso esercitazioni pratiche al fine di far acquisire esperienza ai discenti) ed in servizi di consulenza forniti ad altre imprese, operanti nei più svariati settori.

La formazione in aula è svolta in una struttura didattica accreditata presso la Regione Campania per lo svolgimento di attività di formazione professionale, situata a Napoli in Via A. D'Alessandro 94, che rappresenta la sede sia legale che operativa dell'ARES.

Tale struttura occupa una superficie complessiva netta di circa mq 256 e si estende su un unico piano; è divisa in aule e laboratori quale quello multimediale con 20 postazioni per computer.

Tra i tantissimi corsi offerti si è scelto di gestire con le tecniche in precedenza esposte, un corso SAP (strettamente collegato al settore informatico), cioè un corso dove il vincolo di capacità sulle risorse che si hanno a disposizione è particolarmente importante in quanto essendo questo un corso che si svolge in aula tramite l'ausilio del computer è fondamentale avere esattamente a disposizione tante risorse quanti sono i discenti.

Il SAP (Systems Applications and Products for data processing) è un sistema informativo che consente alle organizzazioni di gestire le risorse interne all'azienda e di integrare completamente tutti i sistemi produttivi. Il sistema SAP è un multipiattaforma, multibase, modulare ed è aperto alle integrazioni. Il software consta dei numerosi moduli, tra cui i principali sono: Financial Accounting (FI-CO); Human Resources (HR); Manufacturing and Logistics (MM); Sales and Distribution (SD).

L'ARES organizza per ognuno dei suddetti moduli, e per altri ancora, dei corsi specifici, con durata, prezzo e programma differenti per cui una volta deciso di implementare lo yield ad un corso SAP si è dovuta effettuare un'ulteriore decisione tra i differenti corsi che l'associazione svolge, che, nello specifico sono:

- Overview di SAP/R3 corso della durata di 16 ore che ha l'obiettivo di formare figure professionali in grado di interfacciare un ampio spettro di data entry ed ha lo scopo di preparare il discente alla fruizione dei moduli successivi;
- Corso SAP/R3 modulo MM della durata di 40 ore in cui vengono approfondite le applicazioni della piattaforma Sap R/3 sia per gestire i processi di approvvigionamento di beni e servizi che per tutte le esigenze di vendita, spedizione e fatturazione;
- Corso SAP/R3 modulo CO della durata di 40 ore il cui obiettivo è la preparazione dei discenti alla gestione, tramite piattaforma Sap R/3, della contabilità generale ed analitica ed il controllo della gestione aziendale da un punto di vista passivo;
- Corso SAP/R3 modulo FI della durata di 40 ore il cui scopo è la gestione, tramite piattaforma Sap R/3, della contabilità generale ed analitica ed il controllo della gestione aziendale da un punto di vista interno;

- Corso SAP/R3 modulo HR della durata di 40 ore che approfondisce la gestione, tramite piattaforma Sap R/3, del personale dell'azienda;
- Corso ABAP4 della durata di 40 ore che consente di approfondire il linguaggio di programmazione ABAP4 per SAP/R3.

Tra i vari corsi si è scelto l'ultimo corso, cioè il corso ABAP4 per SAP/R3.

La scelta è stata completamente arbitraria in quanto tutti i corsi sono egualmente validi per l'implementazione dello yield management in quanto sono accomunati dal medesimo obiettivo: gestire nel modo migliore possibile le risorse "postazioni computer" in numero limitato e trarre da esse il massimo profitto.

Il corso ABAP4 per SAP R/3 si tiene nella sede dell'ARES in Via A. D'Alessandro in un'aula attrezzata con 20 PC di ultima generazione interamente dedicati alla programmazione ABAP in ambiente SAP R/3 versione 4.6B.

Il corso si svolge direttamente in laboratorio e prevede, per ogni giorno la spiegazione prima teorica e poi pratica dell'argomento, in modo da poter provare immediatamente dopo la lezione teorica quanto appreso. La durata del corso è di 40 ore distribuite in 8 ore giornaliere per i cinque giorni lavorativi settimanali; si tratta quindi di un corso full-immersion.

Nelle 40 ore sono trattati tutti gli argomenti previsti nel programma; in particolare, ogni giorno si trattano due argomenti, uno nella lezione della mattina ed uno per la lezione del pomeriggio e per ognuno, dopo la spiegazione teorica dell'argomento, si procede ad una lezione pratica molto approfondita nella quale i docenti seguono personalmente i discenti e aiutano questi ultimi ad applicare al computer quanto studiato.

Il corso è autofinanziato dall'Ares, cioè la Regione Campania non versa alcun contributo per la realizzazione dello stesso; il suo costo è di 300 €, comprensivi di IVA, da pagarsi eventualmente anche a rate, di cui la prima è corrisposta solo dopo la fruizione del primo giorno di lezione.

Per partecipare al corso occorre accedere al sito www.infoares.it e registrarsi per il corso specifico; la registrazione non è impegnativa per cui può anche accadere che, l'utente, pur essendosi registrato, non si presenti al corso alla data stabilita. E' proprio questa eventualità, non troppo remota, che si tenta di gestire e limitare praticando l'overbooking. Infatti, non essendo la prenotazione in alcun modo vincolante, spesso l'Ares si è trovata tenere dei corsi con 15/16 discenti, perdendo una quota, anche abbastanza sostanziosa, di potenziali profitti.

Come anticipato, il processo di fruizione dei servizi che l'Ares offre si snoda in due parti, fisicamente e temporalmente distinte:

- 1) Prenotazione del Servizio;

2) Acquisto del servizio (pagamento) e fruizione del servizio.

che graficamente possono essere rappresentate come segue:



Figura 4.16: Processo tipico dell'Ares

E' proprio la differenza temporale tra la fase di prenotazione e quella di acquisto del servizio che potrebbe procurare danni economici all'Ares in quanto, se si verificano dei no-show si rischia di non saturare completamente la capacità di cui si dispone e di perdere dei potenziali profitti; pertanto queste fasi devono essere gestite nel migliore dei modi attraverso opportune politiche di overbooking.

Se, invece, la fase di prenotazione e quella di acquisto avvenissero contemporaneamente non ci sarebbero questi problemi in quanto anche se avvenissero dei no-show questi ultimi non porterebbero alcun danno economico all'azienda in quanto il valore del servizio prenotato è stato già corrisposto interamente precedentemente.

4.11.2 Implementazione del metodo

L'implementazione dello yield management è stata svolta parallelamente con due metodi alternativi entrambi volti e incentrati alla massimizzazione del profitto dell'Ares tramite un'adeguata gestione integrata della domanda del servizio e della capacità dello stesso.

In ogni caso si è cercato di individuare il limite ottimo di overbooking attraverso l'ausilio due semplici tecniche gestionali:

- l'analisi sui valori medi;
- l'analisi tramite foglio di calcolo.

L'obiettivo di queste tecniche risulta il medesimo e si sostanzia nella mediazione tra due esigenze differenti e contrapposte: tutelarsi da eventuali no-show del corso accettando prenotazioni dei posti in

numero maggiore di quanti effettivamente se ne dispone ed, al contempo, non aumentare eccessivamente questo limite in quanto occorre corrispondere un rimborso economico nel caso in cui un cliente prenotato non abbia un posto riservato.

L'implementazione ha richiesto diversi dati che sono stati ricavati dai record storici di cui l'Ares è dotato.

In particolare sono stati necessari degli studi per ricavare:

- la probabilità che un cliente prenotato effettivamente si presenti per seguire il corso;
- la probabilità da associare a zero no-show, ad un no-show, a due no-show e così via fino a sette, caso massimo di no-show presentatosi in tutta la storia dell'Ares;
- il costo che l'Ares sostiene non vendendo un posto del corso (Spoilage Cost);
- il costo che l'Ares sostiene quando ad un cliente con prenotazione non è garantito un posto (Walking Cost).

Analisi sui valori medi

L'analisi sui valori medi è la tecnica per la determinazione dell'overbooking più semplice che esista in letteratura.

La determinazione del limite di prenotazioni in overbooking da accettare viene effettuata a partire da due dati in ingresso:

- 1) la capacità di cui l'associazione dispone;
- 2) la probabilità che una volta effettuata la prenotazione il cliente effettivamente si presenti per usufruire del servizio.

Lo schema logico con cui funziona la tecnica è il seguente:

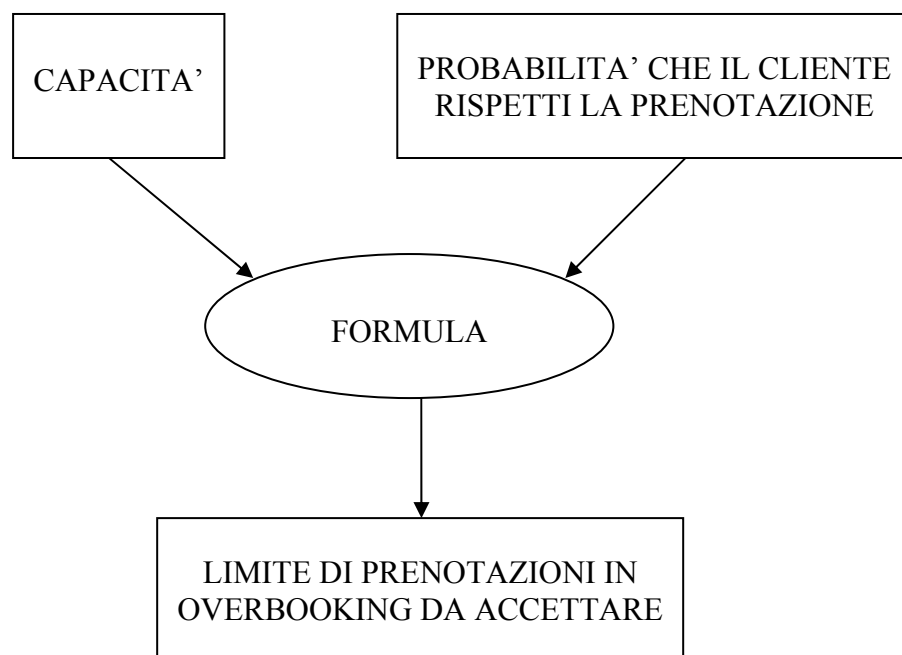


Figura 4.17: Schema di funzionamento dell'analisi sui valori medi

L'input della capacità è chiaramente pari a 20, numero di postazioni computer di cui l'Ares dispone.

La probabilità che il cliente che ha prenotato effettivamente si presenti il giorno stabilito per usufruire del servizio è stata determinata consultando i dati storici di cui l'Ares dispone in abbondanza, operando nel settore da diversi anni ed avendo organizzato e gestito decine e decine di corsi del genere, che necessitano cioè di una postazione computer. Per ognuno dei corsi tenuti in passato si è calcolata la suddetta probabilità e poi è stata fatta una media dei diversi valori di probabilità ottenuti. Il valore finale risultato da quest'analisi è stato 0,81.

Una volta determinati i valori da associare a questi input si è passati alla seconda fase: l'inserimento degli stessi nella formula per determinare il limite di overbooking.

La formula applicata è stata la seguente:

$$BL = C / p$$

dalla quale, sostituendo i valori del caso, risulta che:

$$BL = 20 / 0.78 = \mathbf{25.64}$$

In primo luogo è stato approssimato questo valore che, rappresentando il numero di prenotazioni deve avere necessariamente valore intero. Si è approssimato quindi 25.64 con 26.

Essendo la capacità effettiva pari a 20 è possibile concludere che con questa tecnica il limite di prenotazioni in overbooking da accettare è pari a 6.

Questa tecnica presenta il vantaggio indiscutibile della semplicità, dell'immediatezza e della facilità di implementazione.

Al contempo però presenta due grossi limiti:

- 1) l'approssimazione del numero decimale che sovente si ottiene con un numero intero;
- 2) l'omissione dell'analisi dei costi connessi all'overbooking.

Analisi tramite foglio di calcolo

La scelta del numero ottimo di prenotazioni in overbooking da accettare, in un'analisi più approfondita, non può prescindere dalla valutazione dei costi connessi a tale pratica. Il realtà questa scelta si riduce all'individuazione di un opportuno trade-off tra due tipologie di costi:

- 1) Spoilage cost;
- 2) Walking cost.

Lo spoilage cost rappresenta il costo che l'azienda deve sostenere, in termini di mancato guadagno, quando non riesce a saturare completamente la capacità di cui dispone, in quanto si trova con posti riservati che però al momento dell'erogazione del servizio non sono richiesti da nessuno. Questo costo, che è meglio definibile come mancato profitto, è pari al prezzo unitario a cui si vende il servizio.

Nel caso in esame essendo il prezzo del corso SAP R/3 ABAP4 che l'Ares eroga pari a 300 €, tale risulta lo spoilage cost.

Il walking cost rappresenta il costo che l'azienda sostiene nel momento in cui non riesce a garantire l'erogazione del servizio al cliente pur avendolo, quest'ultimo prenotato precedentemente. Si tratta di un costo di recupero in quanto tenta di ripristinare la fiducia del cliente e l'immagine aziendale che, in modo del tutto comprensibile, diminuiscono quando si verificano situazioni del genere.

L'Ares, qualora si verificassero situazioni del genere offre, in ogni caso, un rimborso spese di 50 € come risarcimento danni per essersi recati alla sede, nel giorno e nell'ora stabilita senza aver potuto usufruire del

servizio, ed inoltre uno sconto del 50% per la fruizione in una data successiva, a scelta discrezionale del cliente, del medesimo corso sul quale si acquisisce diritto di prelazione.

In definitiva il walking cost che l'Ares sostiene è pari a 150 € che rappresenta il mancato guadagno dal corso a seguito dello sconto a cui bisogna aggiungere i 50 € di rimborso spese per cui in totale questo ammonta a 200 €.

Oltre a considerare i costi questo metodo di analisi considera la probabilità che non si verifichino no-show, che se ne verifichi uno e così via fino a considerare il caso massimo di sette no-show, che rappresenta il valore massimo di no-show presentatisi contemporaneamente all'Ares.

Queste probabilità hanno valori variabili da settore a settore e anche nel medesimo settore tra aziende differenti, e, pertanto, risultano individuabili solo attraverso approfondite analisi dei dati storici aziendali.

Questo metodo, quindi, a differenza del precedente, richiede una serie più numerosa di dati in input, la cui ricerca di alcuni risulta abbastanza complessa.

Nello specifico il funzionamento del metodo è il seguente:

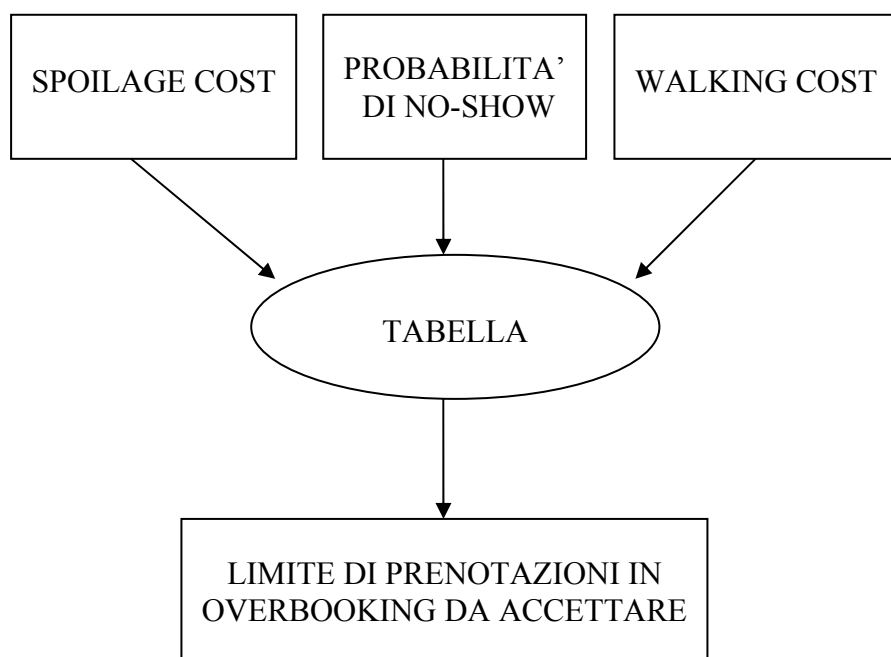


Figura 4.18: Schema di funzionamento dell'analisi tramite foglio di calcolo

Una volta che si dispone di tutti i dati in input occorre sistamarli in tabella; si inseriscono i no-show con le relative probabilità di accadimento, il valore dello spoilage cost (C_u) e quello del walking cost (C_o).

L'organizzazione dei dati è illustrato nella seguente tabella:

Numero di prenotazioni in overbooking									
No-show	Probabilità	0	1	2	3	4	5	6	7
0	xxx	0	C_o	$2 \cdot C_o$	$3 \cdot C_o$	$4 \cdot C_o$	$5 \cdot C_o$	$6 \cdot C_o$	$7 \cdot C_o$
1	xxx	C_u	0	C_o	$2 \cdot C_o$	$3 \cdot C_o$	$4 \cdot C_o$	$5 \cdot C_o$	$6 \cdot C_o$
2	xxx	$2 \cdot C_u$	C_u	0	C_o	$2 \cdot C_o$	$3 \cdot C_o$	$4 \cdot C_o$	$5 \cdot C_o$
3	xxx	$3 \cdot C_u$	$2 \cdot C_u$	C_u	0	C_o	$2 \cdot C_o$	$3 \cdot C_o$	$4 \cdot C_o$
4	xxx	$4 \cdot C_u$	$3 \cdot C_u$	$2 \cdot C_u$	C_u	0	C_o	$2 \cdot C_o$	$3 \cdot C_o$
5	xxx	$5 \cdot C_u$	$4 \cdot C_u$	$3 \cdot C_u$	$2 \cdot C_u$	C_u	0	C_o	$2 \cdot C_o$
6	xxx	$6 \cdot C_u$	$5 \cdot C_u$	$4 \cdot C_u$	$3 \cdot C_u$	$2 \cdot C_u$	C_u	0	C_o
7	xxx	$7 \cdot C_u$	$6 \cdot C_u$	$5 \cdot C_u$	$4 \cdot C_u$	$3 \cdot C_u$	$2 \cdot C_u$	C_u	0
Costo totale									

Tabella 4.2: Organizzazione dei dati in tabella

Sostituendo i dati relativi all'Ares si ottiene:

Numero di prenotazioni in overbooking									
No-show	Probabilità	0	1	2	3	4	5	6	7
0	0.05	0	200	400	600	800	1000	1200	1400
1	0.10	300	0	200	400	600	800	1000	1200
2	0.20	600	300	0	200	400	600	800	1000
3	0.20	900	600	300	0	200	400	600	800
4	0.25	1200	900	600	300	0	200	400	600
5	0.10	1500	1200	900	600	300	0	200	400
6	0.05	1800	1500	1200	900	600	300	0	200
7	0.05	2100	1800	1500	1200	900	600	300	0
Costo totale		975	700	475	365	325	425	575	750

Tabella 4.3: Tabella nel caso dell'Ares

Una volta inseriti i dati in tabella occorre calcolare l'ultima riga della stessa; per ricavarla bisogna, per ogni colonna, moltiplicare ogni suo costo per la relativa probabilità di accadimento e sommare i vari costi ottenuti, occorre poi iterare il procedimento per tutte le colonne della tabella.

Nel momento in cui si dispone dei costi attesi totali relativi a tutte le colonne occorre tra questi scegliere quello che ha valore minimo. Una volta individuato il costo medio minimo occorre risalire verticalmente alla cima della medesima colonna e ricavare il numero di prenotazioni in overbooking corrispondente.

Applicando questa tecnica come illustrato finora si è giunti alla seguente conclusione: accettare 4 prenotazioni in overbooking.

Confronto tra le tecniche

I risultati ottenuti implementando le due tecniche sono differenti: applicando la tecnica di analisi sui valori medi risulta che il numero ottimo di prenotazioni in overbooking da accettare è pari a 6 mentre implementando la tecnica di analisi tramite foglio di calcolo si individua il valore ottimo in 4.

La differenza tra i risultati ottenuti con le due tecniche è certamente sensibile, anche se apparentemente può sembrare il contrario. Infatti, considerando non i numeri in modo assoluto ma in relazione alla capacità che occorre gestire la prima propone di accettare il 30% di prenotazioni in più rispetto a quelle che si potrebbero con certezza servire mentre la seconda si mantiene più cauta e individua un limite di booking pari al 20% della capacità effettiva.

In definitiva i numeri proposti, 6 e 4, considerati semplicemente come tali sembrano apparentemente molto simili ma se, come detto, si considerano in relazione a grosse capacità possono portare a risultati totalmente differenti; si pensi al settore aereo dove si gestiscono oltre 200 risorse o il settore alberghiero in cui le camere talvolta possono arrivare anche a mille è chiaro ad esempio, per quest'ultimo caso che gestire 200 o 300 camere in overbooking può portare a risultati economici totalmente differenti.

La differenza tra i numeri ottenuti con le due tecniche è spiegabile considerando che mentre la prima considera semplicemente la capacità e la probabilità che il cliente mantenga la prenotazione la seconda considera anche i costi; per tale ragione è chiaro comprendere che la prima tende a sovrastimare il valore ottimo che, non considerando i costi che l'overbooking comporta, si potrebbe portare, al limite, anche all'infinito, assicurando all'azienda la completa saturazione della capacità in ogni situazione. Al contrario, considerando anche i costi la situazione cambia radicalmente in quanto il ragionamento precedente non ha più valore poiché occorre considerare anche gli esborsi economici che l'azienda sostiene ogni volta che si trova in overbooking: è chiaro che il valore ottimo deve risultare contenuto e realistico, bilanciando opportunamente il costo opportunità connesso alla saturazione della capacità con il costo che qualora si sovrassaturi quest'ultima occorre corrispondere ai clienti.

In conclusione è possibile affermare che il metodo dell'analisi tramite foglio di calcolo supera entrambi i limiti dell'analisi sui valori medi presentata in quanto fornisce direttamente un numero intero di prenotazioni in overbooking da accettare ed inoltre considera i costi relativi all'overbooking; si tratta,

quindi, indubbiamente, di un metodo più completo in quanto considera l'elemento chiave rispetto al quale impostare qualsiasi scelta gestionale: il costo.

La scelta ottima è effettuata con maggiore cognizione di causa e effettuando un'analisi a tutto tondo dello scenario di riferimento, considerando delle variabili che risultano irrinunciabili per una scelta del genere. A fronte quindi di un piccolo incremento di difficoltà dal punto di vista implementativo e di ricerca dei dati la tecnica di analisi tramite foglio di calcolo conduce, senza dubbio, ad una scelta più valida e migliore di quella ottenibile attraverso l'analisi sui valori medi.

4.12 Gestione dell'overbooking attraverso un software implementato in Microsoft Access

L'architettura generale del sistema è implementata in ambiente Microsoft Windows su piattaforma Microsoft Access 2007. Esso può essere utilizzato da qualsiasi operatore per gestire le prenotazioni che giungono alla struttura.

Alla base del software vi è una procedura step to step (passo passo) che permette di inserire una prenotazione con tutte le caratteristiche: Dati del cliente, date prenotazione, tipologia trattamento (tipo di corso, scoutistiche, ecc). Essa si esterna attraverso un'interfaccia grafica, progettata per rendere rapido ed intuitivo l'accesso a tutte le funzioni del programma:

- Verifica in tempo reale della disponibilità e dei prezzi;
- L'accesso alle operazioni più frequenti tramite menù a tendina e bottoni grafici consente all'utente di dare risposte immediate ai propri clienti;
- Gestione giornaliera delle prenotazioni in overbooking;
- Cancellazione automatica delle prenotazioni;
- Modifica delle prenotazioni;
- Guida alla fase di check-in e check-out;
- Archivio storico clienti e prenotazioni.

Le principali funzionalità del software sono riportate nei seguenti diagrammi di flusso

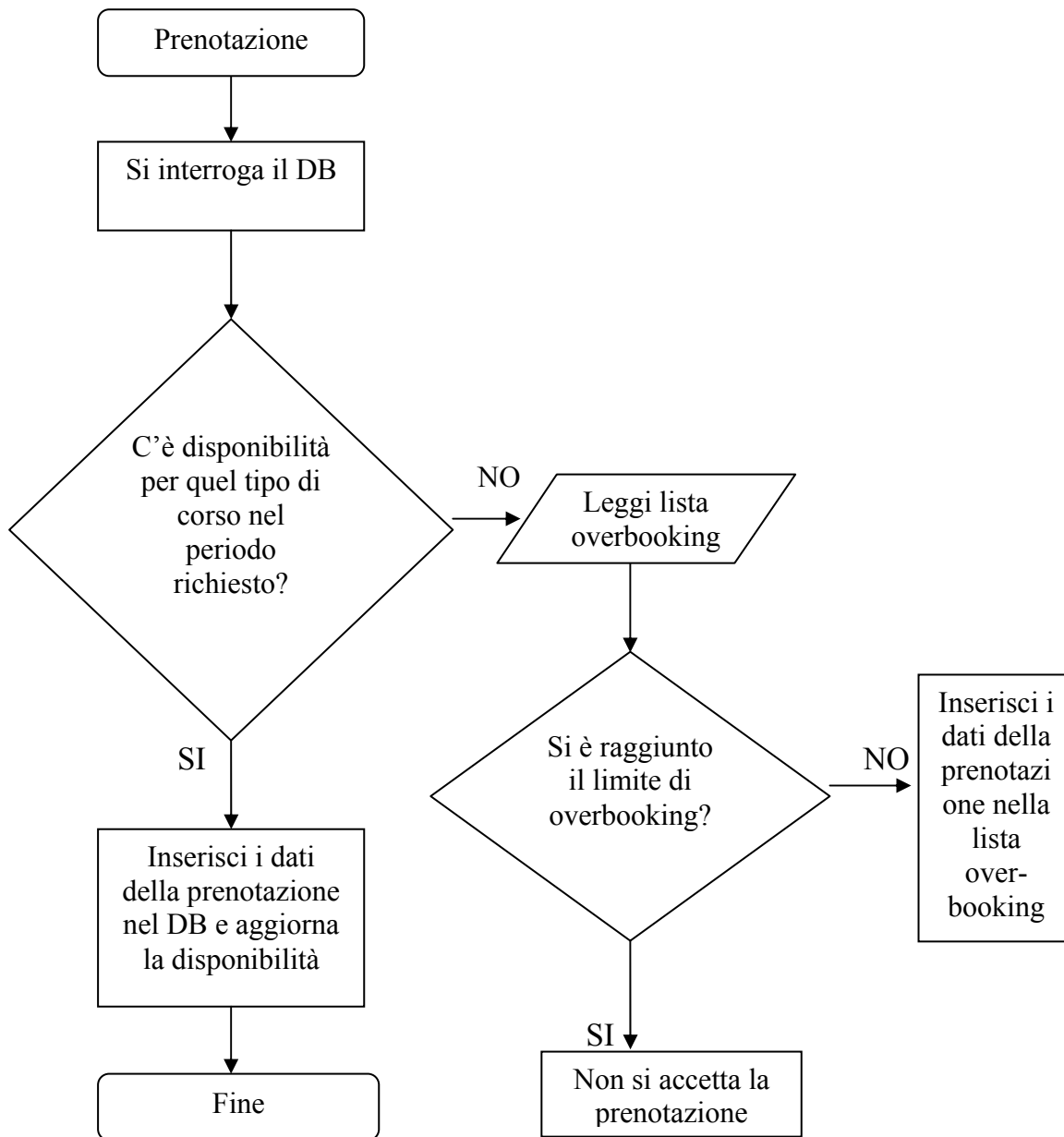
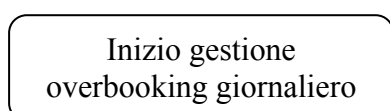


Fig.4.19 Flow chart: fase “prenotazione”

Il secondo flow chart, fig.4.20, è riferito alla fase più importante e delicata di un processo di booking: la gestione giornaliera delle prenotazioni in overbooking.



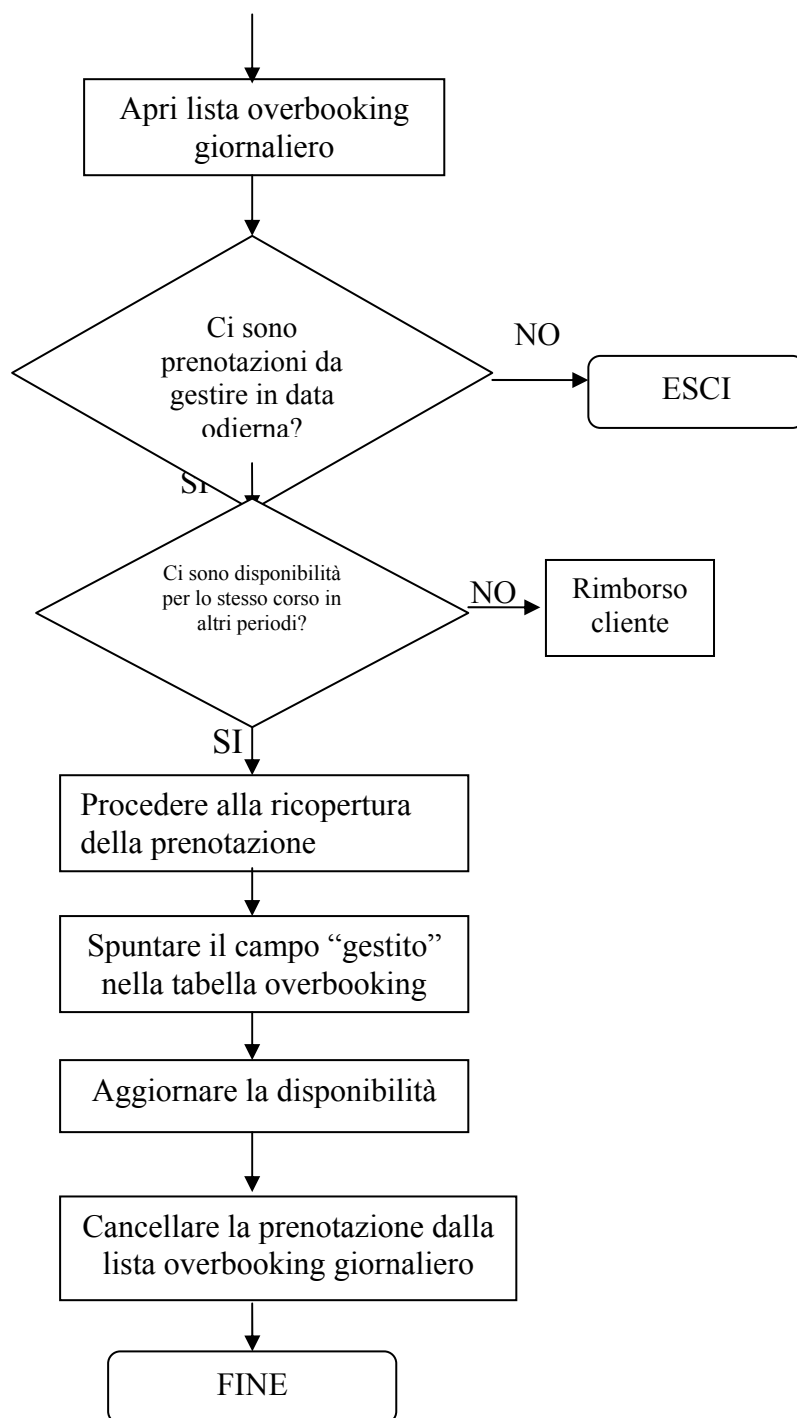


Fig.4.20 flow chart fase: “ gestione overbooking giornaliero”

Una ulteriore fase prevista dal software è quella relativa alla cancellazione o modifica di una prenotazione. Il flusso logico previsto per la cancellazione è riportato in fig. 4.21.

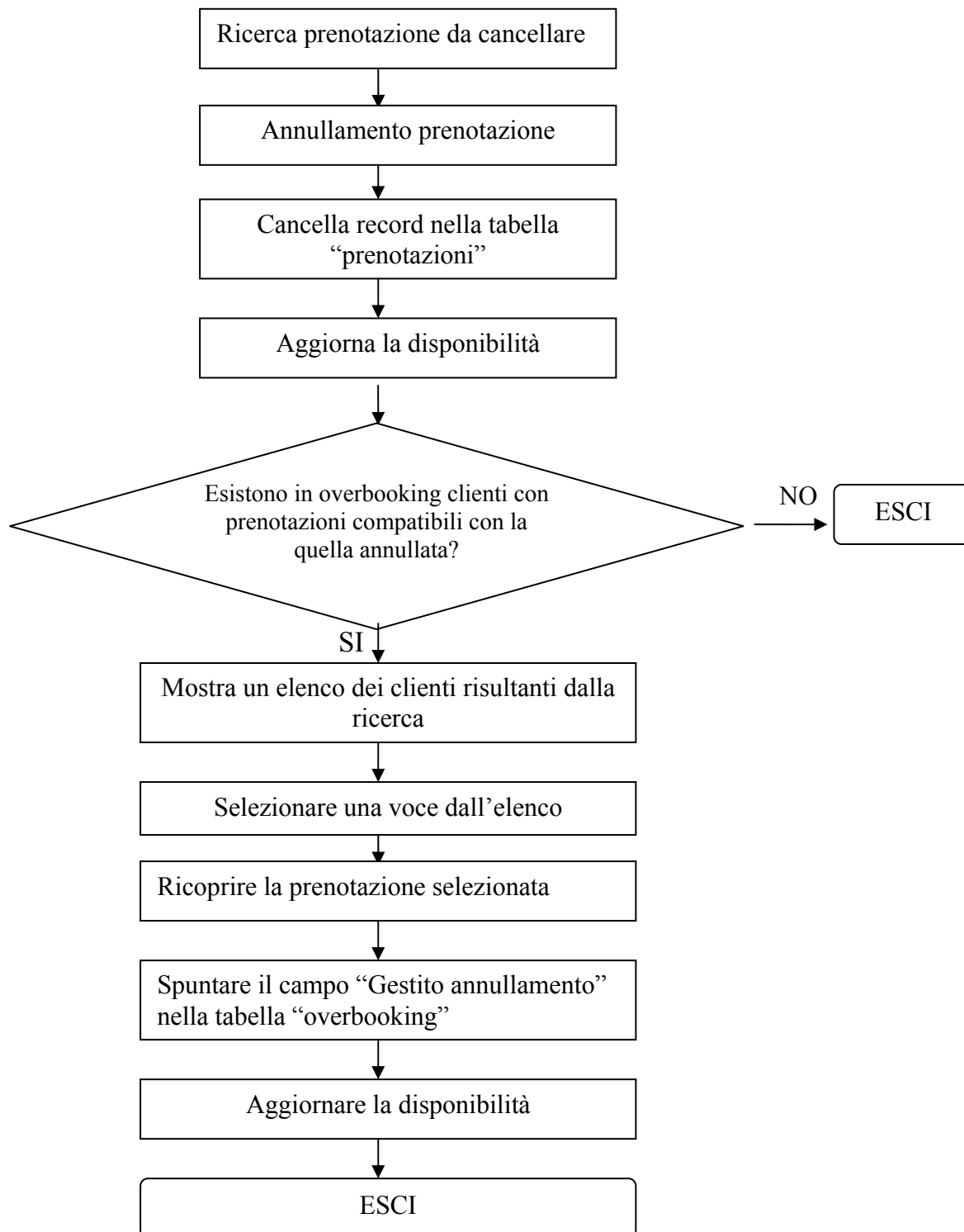
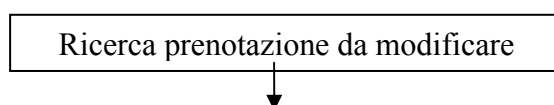


Figura 4.21 flow chart fase “Annullamento prenotazione”

Il flusso logico, invece, previsto per la cancellazione è riportato in fig. 3.5.



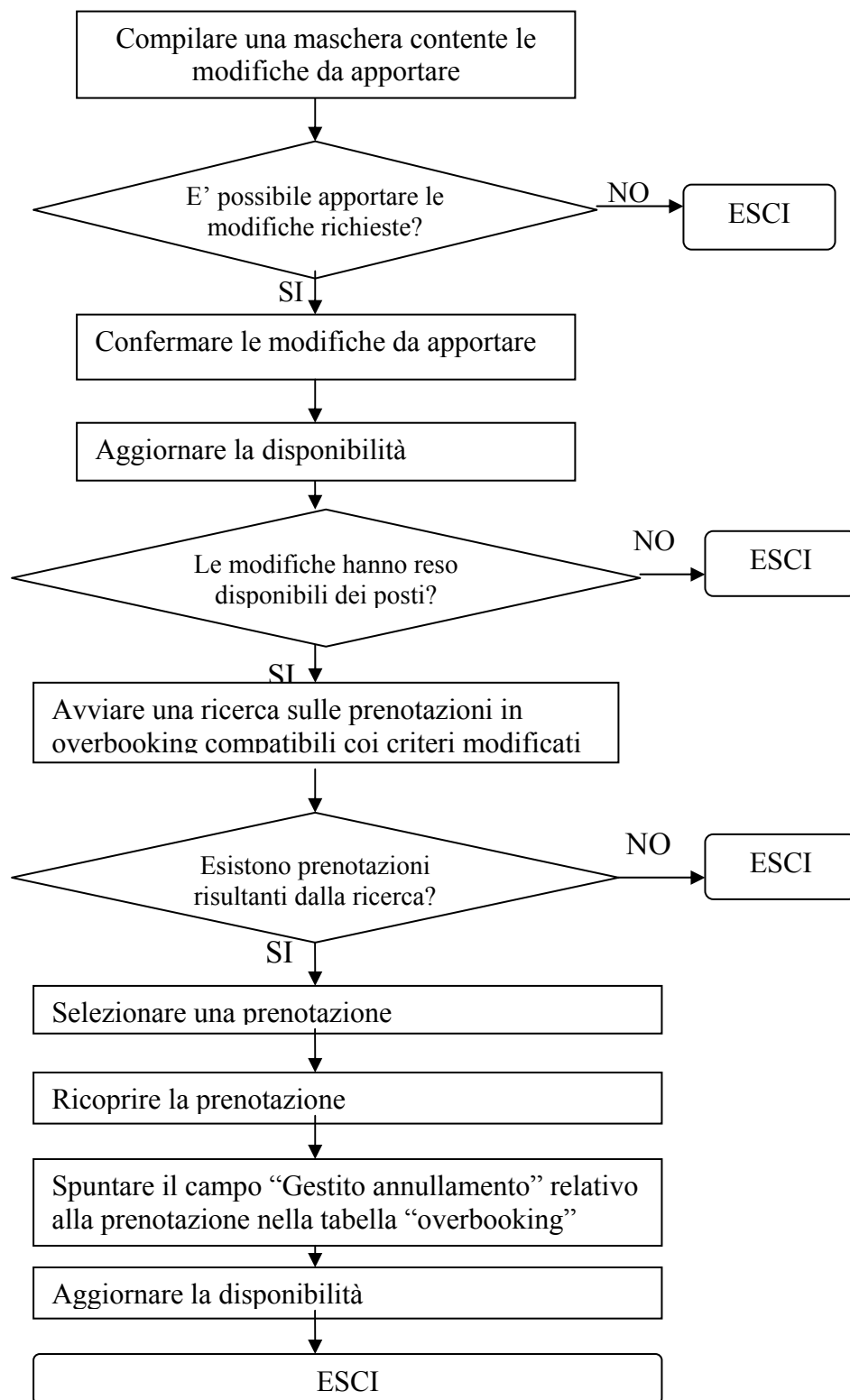
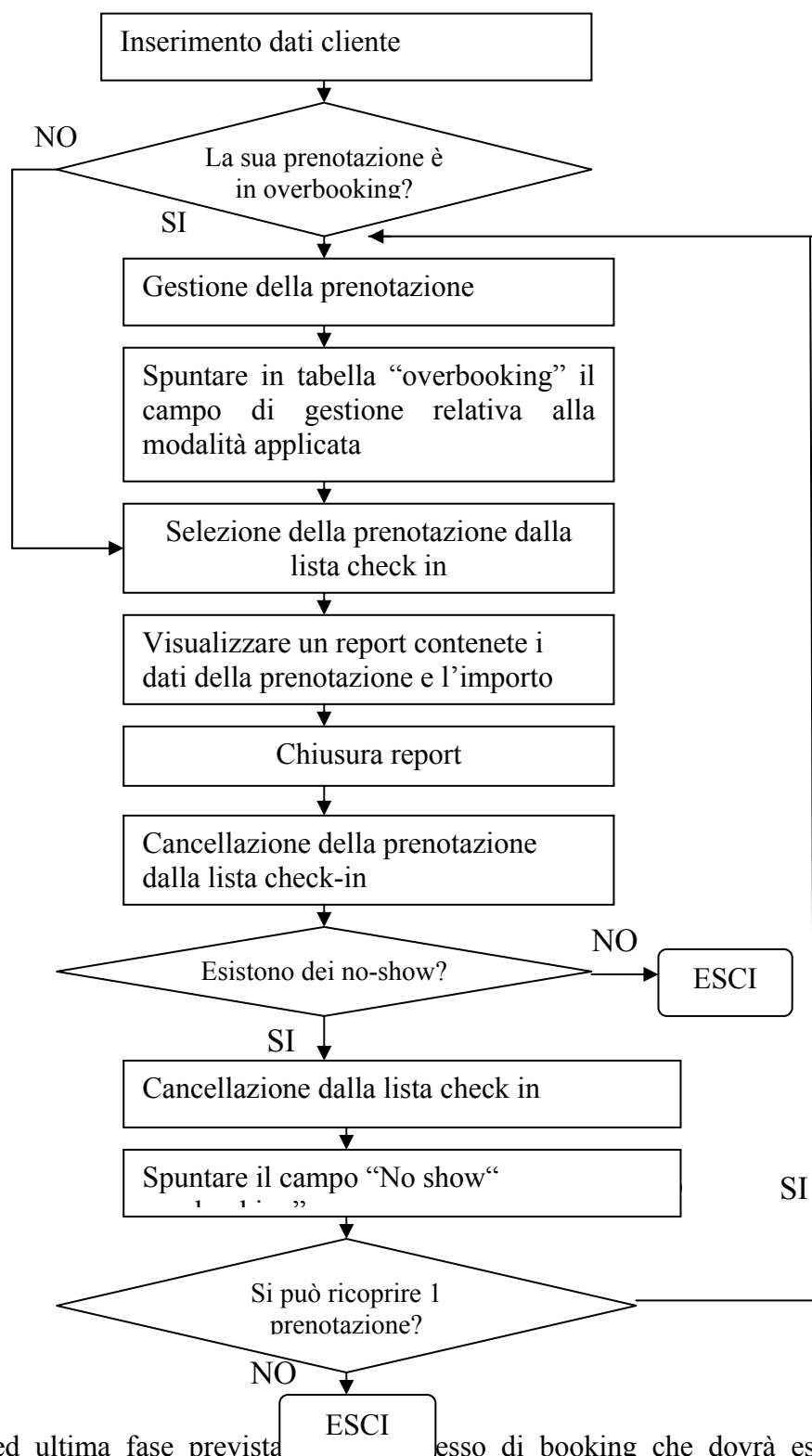


Figura 4.22 flow chart fase “Modifica prenotazione”

La quarta fase prevista dal software è relativa al check in. Essa consentirà, all’operatore di prendere visione dei dati del cliente il primo giorno di corso relativamente alla sua prenotazione.



La quinta ed ultima fase prevista è il processo di booking che dovrà essere gestito dal software

Figura 4.23 Flow chart: fase: “check in”

Il cliente ha terminato un corso e gli si deve rilasciare la fattura. Il flusso logico della suddetta fase è riportato in fig.4.24.

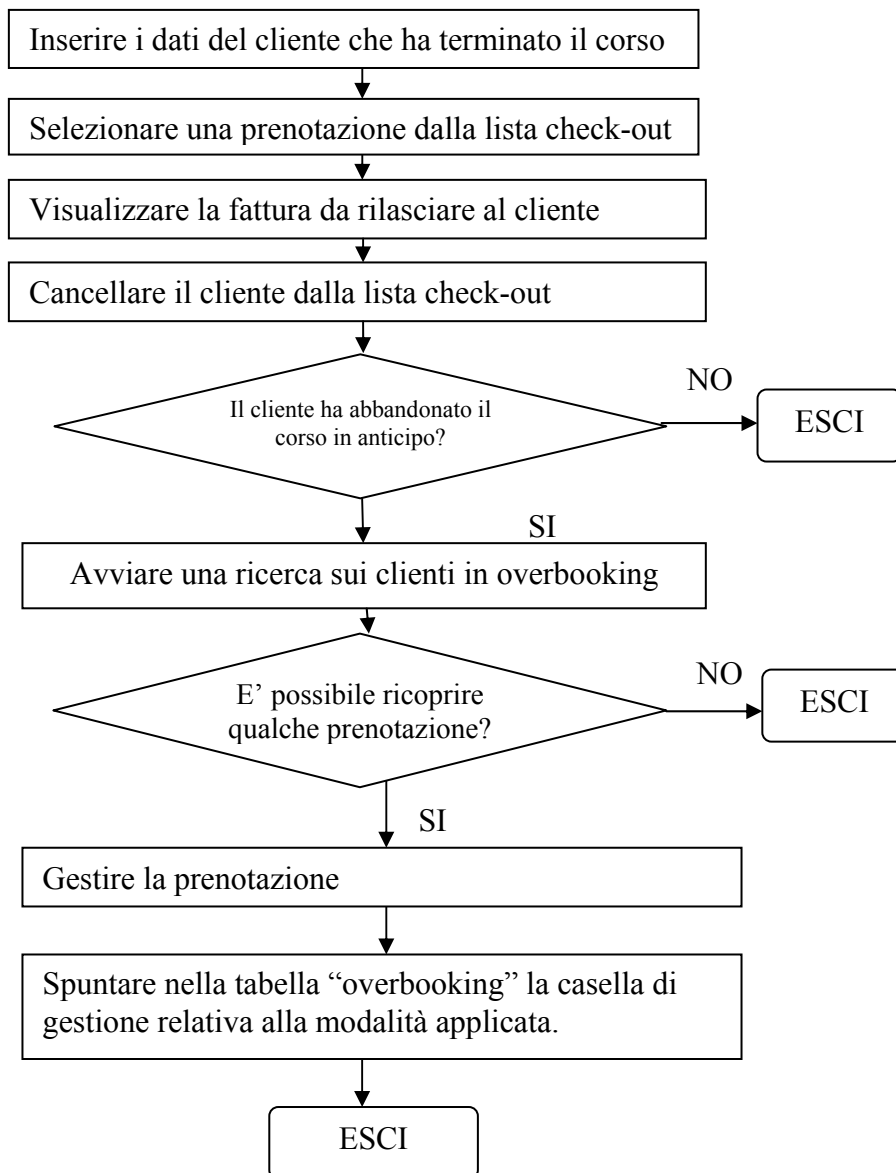


Fig.4.24 Flow chart: fase: “check out”

4.12.1 Come funziona il software

L'accesso al software si configura attraverso la schermata iniziale "Pannello di controllo", riportata in fig. 3.8 e composta da cinque bottoni:

- il primo "prenotazioni" viene cliccato all'atto di una nuova prenotazione pervenuta o attraverso chiamata o via fax o via mail;
- il secondo "gestione overbooking", viene cliccato ogni giorno dall'utente per prendere visione delle persone che quotidianamente sono in overbooking e cercare, quindi di gestire le loro prenotazioni.
- Il terzo "prenotazioni cancellate" viene cliccato quando all'ARES viene data comunicazione della disdetta o della volontà di un cliente di modificare la sua prenotazione. Il limite massimo entro cui la rinuncia può avvenire è 24 ore dalla fase di accettazione.
- Il quarto bottone: "check in", viene cliccato ogni qualvolta un cliente giunge nella struttura. Inserendo i dati del cliente l'utente prende visione della prenotazione effettuata da quest'ultimo, dell'importo che deve pagare e del numero della chiave magnetica relativa alla sua stanza.
- Il quinto "Check out" verrà cliccato dall'utente quando un cliente, terminato il corso, si presenta alla segreteria organizzativa per la fattura ed eventualmente l'attestato finale.



Fig. 4.25 Pannello di controllo

Prima fase: prenotazioni

La prima fase che si andrà ad analizzare sarà quella relativa alla gestione delle prenotazioni. Quando perviene una prenotazione per un corso, l'utente cliccando sul bottone "prenotazione", dal pannello di controllo, potrà accedere alla finestra "Raccolta dati prenotazione", all'interno della quale inserire i dati relativi alla prenotazione:

- **Tipo di corso**
- **Dal:**

- **Al**
- **Modalità**

Terminato l’inserimento di tali parametri si potrà cliccare il bottone “Ricerca disponibilità”, fig. 4.26, posto in basso a destra per avviare la ricerca dei posti disponibili e qualora ve ne siano quindi, terminare la prenotazione.



Figura 4.26 Bottone “Ricerca disponibilità”

In caso di errato inserimento di dati o di errore iniziale della maschera “Raccolta dati prenotazioni”, si potrà tornare al Pannello di controllo per avviare una nuova azione se lo si desidera, cliccando sul bottone posto in basso a sinistra “Torna al pannello di controllo”, fig 4.27.



Figura 4.27 Bottone “Torna al pannello di controllo”

Qualora per distrazione o qualsiasi altra causa, si clicchi il bottone “Ricerca disponibilità” senza aver opportunamente compilato i campi, il software provvederà a rendere tale tasto inattivo ed a lanciare una serie di messaggi di errore, quali:

2) *Inserisci le date* fig. 4.28: tale messaggio verrà visualizzato quando o il campo “dal” o il campo “al” non saranno stati compilati.

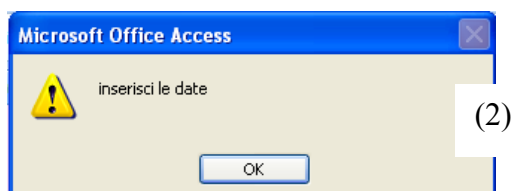


Figura 4.28 messaggio errore

3) *Inserisci il trattamento*, fig 4.29: tale messaggio verrà visualizzato quando il campo “trattamento” non è stato compilato. Ai fini della compilazione sarà obbligatorio scegliere una delle voci presenti nel menù a tendina; non sono ammesse modifiche a tale menù.

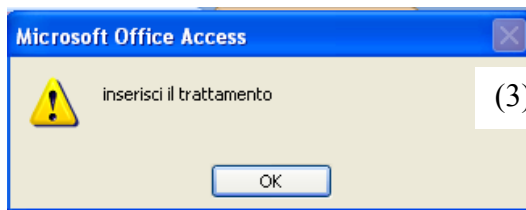


Figura 4.29 messaggio errore

4) *La data richiesta non è valida* fig 4.30: è un messaggio che avverte l'utente che la data inserita è precedente a quella in cui si sta effettuando la prenotazione.

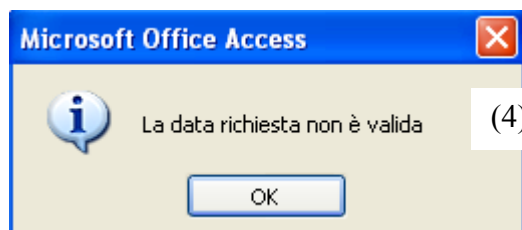


Figura 4.30 messaggio errore

5) *Le date non sono in ordine cronologico*, fig 4.31: il seguente messaggio verrà visualizzato quando si inserirà una data nel campo "Al" antecedente a quella inserita nel campo "Dal".

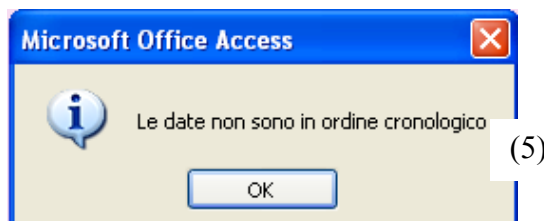


Figura 4.31 messaggio errore

Tali messaggi rendono l'utilizzo del software particolarmente semplice perché guidano l'utente in ogni sua azione.

Una volta compilata correttamente la maschera "Raccolta parametri", cliccando sul bottone posto in basso, verrà avviata una fase di ricerca dei corsi disponibili e dei relativi posti ed automaticamente si aprirà una nuova finestra in cui verranno visualizzati i risultati di tale ricerca.

E' possibile visualizzare in questa nuova maschera una serie di campi, alcuni dei quali, come "Tipo di corso", "Dal", "Al", "Trattamento", già compilati perché richiamano i dati inseriti nella sezione

precedente, altri quali: “Nome”, “Cognome”, “Documento”, “Telefono”, “Cellulare” ed “Email”, che dovranno essere riempiti con i dati del cliente che sta effettuando la prenotazione.

Particolare menzione spetta ai campi “Importo” ed “Acconto”, i quali verranno automaticamente compilati, in quanto il software effettua un calcolo sulla base della politica tariffaria effettuata dall’ente di formazione e del numero di giorni relativi alla prenotazione effettuata. Il primo sarà quindi indicativo della somma che il cliente dovrà pagare se darà conferma della sua prenotazione. Il secondo sarà pari al 30% dell’importo e dovrà essere versato dal cliente subito dopo la prenotazione come garanzia data all’ente, affinché gli venga riservato un posto all’interno del corso di formazione pescelto.

Si può notare inoltre all’interno della finestra, nella sua parte superiore, la presenza di altri due campi:

- “Suggerimento”: esso verrà automaticamente compilato dal software. Quest’ultimo è stato infatti progettato per ottimizzare la disposizione dei posti, così da prevenire ,almeno in parte, il fenomeno dell’overbooking.
- “Posti disponibili”: cliccando sulla freccia posta a destra della casella di testo si aprirà un menù a tendina all’interno del quale verranno visualizzate tutti i posti relativi ai corsi disponibili nel periodo selezionato.

E’ possibile infine notare nella parte inferiore della maschera “Prenotazione”, tre bottoni, di cui quello a destra, denominato “Prenota”, servirà a completare la fase di prenotazione qualora vi siano posti disponibili e tutti i campi siano stati opportunamente compilati. Il bottone al centro, “Prenota in lista overbooking”, servirà ad inserire i dati relativi alla prenotazione nella lista overbooking qualora non vi siano posti liberi nel periodo richiesto. Il bottone “Torna a raccolta parametri”, posto sulla sinistra, consentirà all’utente di ritornare alla maschera precedente qualora voglia modificare uno o più parametri inseriti.

Seconda Fase: Gestione overbooking giornaliero

La seconda fase che si andrà ad analizzare sarà quella relativa alla gestione giornaliera delle prenotazioni inserite nella lista overbooking. Un primo approccio che si è pensato di seguire per gestire l’overbooking, consiste nel ricercare all’interno del database la disponibilità di posti di tipologia simile rispetto a quella prenotata dal cliente. Quest’ultimo ovviamente pagherà un prezzo relativo alla tipologia di corso da lui prenotata a prezzo scontato. In quest’ottica è facilmente comprensibile che l’ente subirà una perdita economica pari alla differenza tra il prezzo pieno e prezzo scontato. A fronte di tale perdita, occorre però sottolineare il ricavo, non in termini economici, bensì relativi alla mancata perdita di immagine.

A tal fine, quotidianamente l'utente dovrà cliccare sul bottone "Gestione overbooking" dal pannello di controllo, fig. 3.27, per visualizzare se ci sono ed eventualmente quali e quante saranno le prenotazioni che egli dovrà ricoprire nell'arco della giornata.

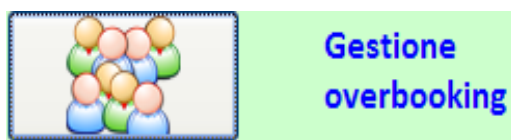


Figura 4.32 Bottone "Gestione overbooking"

Una volta selezionata una delle voci si potrà pigiare sul bottone "Ricerca disponibilità corsi simili", per richiamare la maschera successiva in cui visualizzare se vi siano o meno dei corsi disponibili in quel periodo, dovute magari alla cancellazione di prenotazioni da parte di altri clienti.



Figura 4.33 Bottone "Ricerca disponibilità corsi simili"

Terza fase: Modifica/Cancella prenotazione

La terza fase che si andrà ad analizzare sarà quella relativa alla cancellazione o modifica dei dati relativi ad una prenotazione, effettuata in passato. Quando attraverso fax, mail o chiamata un cliente dà comunicazione all'ente di voler modificare la sua prenotazione o di volerla annullare, l'utente cliccherà sul bottone "Modifica/Cancella Prenotazioni", fig. 4.34.



Figura 4.34 Bottone "Modifica/ cancella prenotazioni"

All'atto di tale pigiatura si aprirà una maschera, "Dati prenotazione da modificare/cancellare", fig. 4.35, all'interno della quale l'utente potrà compilare una serie di campi relativi al cliente: "Nome", "Cognome" ed/od in alternativa alla prenotazione da lui effettuata in passato: "Dal", "Al", così da avviare una ricerca all'interno del database, al termine della quale potrà avere visione della prenotazione da annullare o su cui apportare modifiche.



Figura 4.35 Maschera “ Dati prenotazione da modificare / cancellare”

Qualora l’utente dovesse omettere l’inserimento dei dati all’interno di uno dei campi relativi al cliente o al periodo prenotato, essendo essi obbligatori, o immettere dei dati errati, il software provvederà ad avvertirlo, attraverso la visualizzazione di un messaggio di errore, fig.4.36.

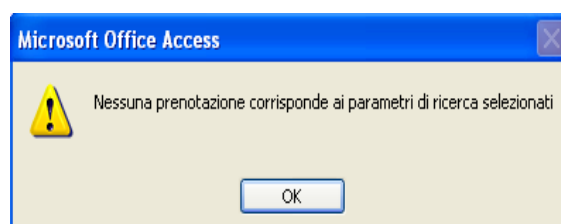


Figura 4.36 messaggio informativo

Una volta inseriti i dati, l’utente potrà cliccare sul bottone “Cerca prenotazioni corrispondenti”, riportato in fig. 4.37, che gli consentirà di visualizzare in una nuova maschera, i risultati della ricerca avviata.



Figura 4.37 Bottone “Cerca prenotazioni corrispondenti”

In caso di apertura non voluta della maschera di fig. 3.43, l’utente potrà tornare al “Pannello di controllo” e decidere di avviare una nuova azione se lo desidera, cliccando sul bottone posto in basso a sinistra “Torna al pannello di controllo”, fig.4.38.



Figura 4.38 bottone “Torna al pannello di controllo”

La maschera all'interno della quale l'utente potrà visualizzare i risultati della sua ricerca, “Elenco prenotazioni ricercate”, sarà composta da due tabelle: “Prenotazioni” ed “Overbooking”, riempite a seconda se la prenotazione ricercata sia in overbooking o meno.

Si può che la tabella “overbooking” differisce da quella “Prenotazioni”, per l'assenza di un campo: “Num. Di corso”; questo a voler sottolineare che la prenotazione è stata immessa nella lista overbooking proprio perchè non si è riusciti ad associare ad essa una posto al corso scelto perchè completo.

Quarta fase : Check in

Il flusso di azioni relative alla fase di check-in ha inizio ogni qualvolta un cliente giunge nella struttura per usufruire del servizio oggetto di una prenotazione fatta in precedenza. Tale flusso verrà innescato cliccando sul bottone “check in”, fig. 3.67 del pannello di controllo.



Figura 4.39 Bottone “check in”

L'operatore in seguito a tale pigiatura, vedrà aprirsi automaticamente una maschera: “Check in”, riportata in fig. 4.40, composta da tre campi: “Documento”, “Nome” e “Cognome”, che dovrà compilare con i dati del cliente appena presentatosi.

Figura 4.40 Maschera “ Check in”

Qualora tale maschera fosse stata erroneamente aperta dall'operatore, egli potrà sempre tornare indietro al pannello di controllo, pigiando sul bottone posto in basso a sinistra, fig. 4.41.



Figura 4.41 Bottone “Torna al pannello di controllo”

Qualora invece, siano stati correttamente compilati tutti i campi della maschera “Check in”, si potrà cliccare sul bottone “Apri lista Check in”, per prendere visione dell'elenco di clienti che in data odierna dovrebbero presentarsi per l'inizio del corso e quindi usufruire del servizio oggetto della loro prenotazione.





Figura 4.42 Bottone “Apri lista check in”

Tale elenco sarà contenuto in una tabella appartenente alla maschera “Lista check in” come può chiaramente vedersi in fig. 4.43.

Lista check in

Nome	Cognome	Documento
lucia	iadiccio	dcclcu






Figura 4.43 Maschera “Lista check in”

Ivi l'operatore dovrà selezionare una delle voci che compaiono in elenco, in particolare quella relativa al cliente che si è presentato previa la visualizzazione di un messaggio di errore che lo inviti a farlo, fig. 4.44.

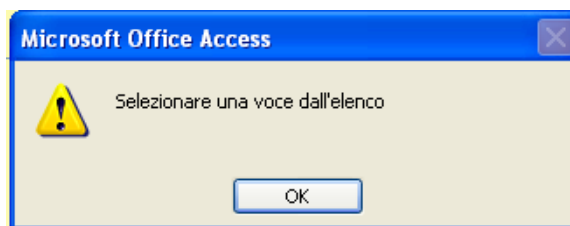


Figura 4.44 messaggio informativo

Se all'interno dell'elenco non dovesse comparire la voce ricercata, potrebbe essere dovuto ad un errore nella compilazione dei dati nella maschera precedente, alla quale si potrà ritornare per apportare le dovute modifiche cliccando sul bottone "Torna a check in", riportato in fig. 4.45.



Figura 4.45 Bottone "Torna a check in"

Se, viceversa l'operatore ha individuato all'interno dell'elenco la voce di suo interesse, dopo averla selezionata, potrà cliccare sul bottone "Apri report check in", che gli consentirà di avere visione di tutti i dati relativi alla prenotazione del cliente. Vedi fig. 4.46.



Figura 4.46 Bottone "Apri report check in"

Tale report riportato, conterrà in particolare tutti i dati relativi al cliente: "Nome", "Cognome", "Documento", alla prenotazione da lui effettuata: "Dal", "Al", "Tipo corso".

In ultimo, tale report conterrà dei campi prettamente riferiti all'aspetto economico del servizio si appresta ad erogare ed il cliente a ricevere: "Importo totale", "Acconto", "acconto dovuto". Il primo si riferisce alla somma di denaro che il cliente deve pagare se vuole usufruire del servizio, il secondo è pari ad un'aliquota dell'importo, che il cliente avrebbe dovuto versare in precedenza come garanzia della sua prenotazione. Tale campo sarà spuntato o meno a seconda che ciò sia o meno accaduto. Infine il terzo, è una differenza tra l'importo totale e l'acconto, qualora quest'ultimo sia stato versato, oppure è pari all'intero acconto in caso contrario.

L'operatore in realtà, potrà imbattersi anche in un report diverso. Ciò accadrà ogni qualvolta esso è riferito ad un cliente la cui prenotazione è stata accettata in overbooking e non la si è riuscita a gestire. L'ente

pertanto è tenuto a rimborsare il cliente: il rimborso previsto dalla legge è pari al doppio dell'acconto versato dal cliente. Si potrà osservare dunque che il campo "Rimborso" sarà pari a zero qualora tale acconto non fosse stato versato e pari a quest'ultimo viceversa.

Il fenomeno dell'overbooking, nasce per far fronte al numero sempre maggiori di no-shows, che possono comportare un mancato utilizzo della risorsa da parte dell'ente e quindi mancati ricavi. Il software è stato implementato oltre che per la gestione dell'intero processo di booking, con particolare riferimento alle tecniche di gestione dell'overbooking, anche per far fronte a tale aspetto.

La maschera si compone infatti di un ulteriore pulsante, oltre ai due precedentemente descritti, "Identifica il cliente come no-show", riportato in fig. 4.47. Esso potrà essere cliccato dall'operatore a fine giornata, così da identificare come no-show tutti coloro che appartengono ancora alla lista check-in (l'elenco si aggiorna quotidianamente) ma che non si sono presentati.



Figura 4.47 Bottone "Identifica il cliente come no-show"

Tale azione consentirà di liberare il posto riservato per tale prenotazione ed avvierà il software ad effettuare una ricerca mirata ad individuare, sulla base di un confronto tra i parametri della prenotazione cancellata e quelli relativi alle prenotazioni inserite nella lista overbooking, quelle più adatte, perché compatibili, a sostituirsi alla prenotazione cancellata. L'operatore verrà informato dell'esistenza di tali prenotazioni attraverso un messaggio, che farà da preludio all'apertura della maschera "Prenotazioni da poter ricoprire", contenente un elenco di tali prenotazioni.

Il flusso di azioni successivo a tale maschera, consistente nella selezione di una voce dell'elenco, nella visualizzazione dei dati relativi alla prenotazione da ricoprire nonché la prenotazione stessa, è uguale a quello descritto nella parte finale della fase terza.

Quinta fase: Check out

La fase di check out può definirsi come l'anello ultimo che chiude il processo di booking. Esso è effettuato per ciascun cliente che seguito l'intero corso di formazione da lui scelto. Tale fase consiste in effetti, nel rilascio da parte dell'albergo di una ricevuta che attesti il servizio "acquistato" dal cliente.

Quando quest'ultimo si presenta alla segreteria organizzativa l'operatore cliccando sul bottone "Check-out", riportato in fig.4.48, vedrà aprirsi una maschera in cui andranno compilati dei campi per l'identificazione del cliente.



Figura 4.48 Bottone "Check out"

I campi contenuti nella suddetta maschera, detta "Check out", saranno in particolare: "Documento", "Nome" e "Cognome", come si può osservare in fig. 4.50.

The image shows a form titled "Check out" in blue text at the top center. Below the title, there are three labels in green text: "Documento:", "Nome:", and "Cognome:". Each label is followed by a white rectangular input field. At the bottom of the form, there are two buttons. The button on the left has a red curved arrow pointing left, and the button on the right has a yellow curved arrow pointing right.

Figura 4.50 Maschera "Check out"

Qualora l'operatore avesse avviato erroneamente la fase di check out, potrà sempre ritornare al pannello di controllo, cliccando sul bottone posto in basso a sinistra e raffigurato in fig. 4.51.



Figura 4.51 Bottone "Torna al pannello di controllo"

Viceversa se si desidera proseguire nella fase di check out, dopo aver compilato correttamente i tre campi con i dati del cliente che sta per abbandonare l'albergo, si potrà pigiare sul bottone riportato in fig. 4.52, per avviare una ricerca su tutti i clienti che presentano dati compatibili con quelli inseriti. Data la presenza del campo "Documento", i casi di omonimia dovrebbero essere pressoché nulli.



Figura 4.52 Bottone “Apri elenco prenotazioni ricercate”

Può verificarsi che la ricerca dia esito nullo, in quanto magari l’operatore abbia sbagliato a digitare i dati relativi al cliente; il software in tal caso provvederà a visualizzare un messaggio di errore che lo avvisi di ciò, fig.4.53.

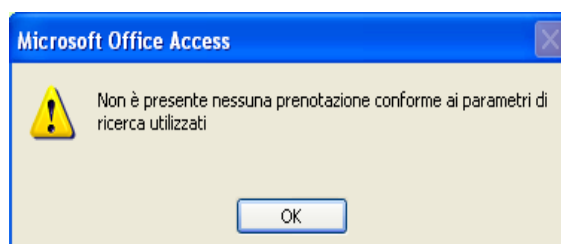


Figura 4.53 messaggio informativo

L’operatore potrà dunque apportare le apposite correzioni. Si aprirà in automatico una nuova maschera: “Elenco prenotazioni ricercate”, contenente il nominativo del cliente ricercato, come riportato in fig. 4.54. Anche in questa maschera sarà data la possibilità all’operatore di tornare a quella precedente, per apportare magari modifiche ai campi compilati, semplicemente cliccando sul bottone posto in basso a sinistra, ”Torna a maschera check out”.



Nome	Cognome	Documento
lucia	iadicicco	dcclcu

Figura 4.54 Maschera “Elenco prenotazioni ricercate”

Dall’elenco di prenotazioni ricercate occorrerà selezionarne una per proseguire nella fase di check out, previa la visualizzazione di un messaggio di errore che inviti a farlo, fig. 4.55.

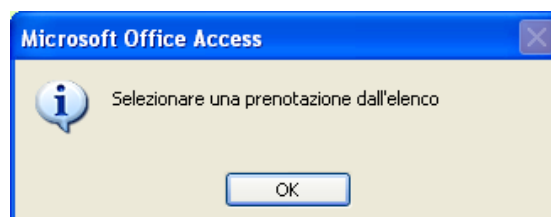


Figura 4.55 messaggio informativo

Selezionata la voce di suo interesse, l'utente potrà cliccare sul bottone posto in basso a destra "Visualizza ricevuta cliente", (vedi fig. 4.56) che consentirà l'apertura di un report contenente la ricevuta da rilasciare al cliente, in cui sono contenuti sia i dati dell'ARES che del servizio "acquistato" dal cliente.

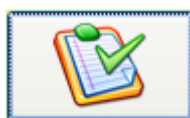


Figura 4.56 Bottone "Visualizza ricevuta cliente"

Tale report, in particolare conterrà nella sua parte superiore un'intestazione indicativa di tutti i dati dell'ente: nome, indirizzo, numero di telefono, partita iva, tipologia di azienda, indirizzo email ed indirizzo web. Successiva all'intestazione vi è una parte centrale contenente i dati del clienti: nome, cognome e documento ed una parte finale in cui è riportato il totale della spesa sostenuta dal cliente, nonché l'acconto con riquadro annesso spuntato o meno a seconda se esso è stato o meno versato.

4.13 Conclusioni

Il lavoro di tesi si è incentrato, oltre sull'individuazione delle principali tecniche utilizzate per saturare la capacità produttiva delle aziende di servizi anche sulla creazione di un software che consentisse di prevenire prima e gestire poi il fenomeno dell'overbooking. In particolare si è voluto effettuare un'analisi economica relativa ai risultati raggiunti gestendo l'overbooking con la politica ritenuta più idonea e quelli che avrebbe potuto ottenere mediante l'applicazione del software implementato.

L'elemento fondamentale a cui ci si è ispirati nella progettazione e realizzazione del programma è stato quello di prevenire il fenomeno dell'overbooking ottimizzando l'assegnazione dei posti liberi relativamente ai corsi di formazione in oggetto. Il software mira infatti ad evitare che in caso di due prenotazioni relative a periodi adiacenti, si assegnino due posti diversi, come spesso è accaduto invece all'ARES. In questo modo si riusciranno a ricoprire due prenotazioni occupando un solo posto. Questa è la politica a cui ci si è ispirati nell'implementazione del software, attraverso la progettazione del campo "Suggerimento". Immettendo tutte le prenotazioni che si sono avute nel periodo compreso tra dicembre

2006 e luglio 2007, si è potuto osservare una prima sostanziale differenza proprio inerente a tale aspetto. Come è facilmente osservabile nel grafico di fig. 4.57, si è riusciti ad effettuare otto prenotazioni attraverso l'ottimizzazione, di cui ben cinque erano state inserite nella lista overbooking.

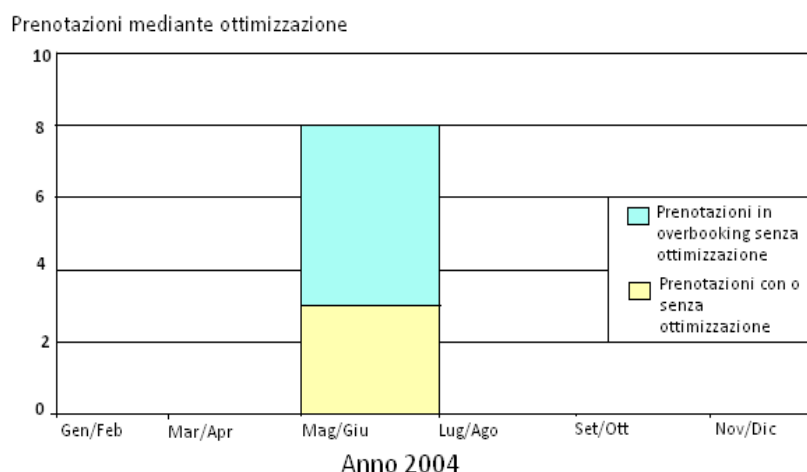


Figura 4.57 Prenotazioni mediante ottimizzazione

Non riuscendo quindi a prevenire il fenomeno dell'overbooking, si dovrà far fronte ad un numero maggiore di prenotazioni da ricoprire. Numero che cresce ulteriormente in virtù del fatto che la politica di gestione dell'ARES non prevede un limite massimo di prenotazioni, superato il quale le successive vengono rifiutate. Tale limite invece, è preso in considerazione dal software che è stato implementato al fine di accettare un numero di prenotazioni in overbooking per ogni tipologia di corso e periodo pari a 4 (vedasi i dati ottenuti dal foglio di calcolo confrontati con i dati dell'analisi marginale). Superato tale limite il programma provvederà attraverso un messaggio a video ad avvertire l'utente che non è più possibile accettare prenotazioni. Dall'analisi dei dati storici, come è rappresentato in fig. 3.89, su circa 70 prenotazioni, si può osservare che ben 20 (rettangolo in rosa) sono quelle in overbooking gestite dall'ARES. Di contro solo 8 (rettangolo arancione) quelle che si sarebbero dovute gestire mediante l'utilizzo del programma, in quanto 7 sono state rifiutate per il raggiungimento del booking limit e le restanti 5 sono state prevenute grazie all'ottimizzazione (gap rappresentato dalla doppia freccia).

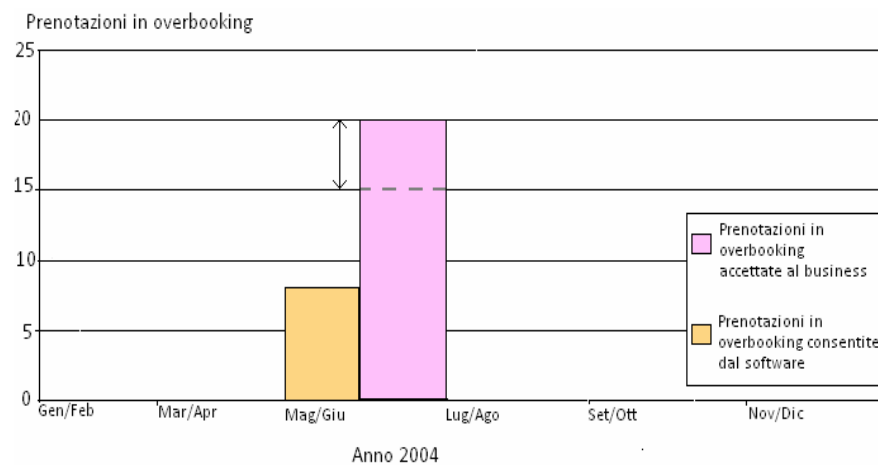


Figura 4.58 prenotazioni in overbooking

INDICE TABELLE

Tabella 1.1 Differenze fondamentali tra sistemi a elevata e a bassa intensità di interazione

Tabella 2.1: Relazione tra tempo di attesa e futuro business

Tabella 3.1. Misure di prestazione di un sistema di servizio

Tabella 3.2. Riepilogo delle misure di prestazione per un sistema M/M/1

Tabella 3.3. Riepilogo delle misure di prestazione per un sistema M/M/s

Tabella 3.4. Confronto tra i modelli con e senza abbandono 50 agenti, $\lambda = 48$, $1/\mu = 1\text{min}$, $1/\theta = 2\text{min}$

Tabella 4.1: Individuazione del valore ottimo di overbooking

Tabella 4.2: Organizzazione dei dati in tabella

Tabella 4.3: Tabella nel caso dell'Ares

INDICE FIGURE

Figura 1.1 Punto di partenza per l'integrazione

Figura 1.2 Il triangolo del servizio

Figura 1.3 Le quattro tipologie di disaccoppiamento

Figura 2.1: Plot di due labor driver in un periodo di pianificazione giornaliero

Figura 2.2: Uso dell'analisi di correlazione

Figura 2.3: Scelta di un intervallo di tempo opportuno per tracciare variabilità di un Driver

Figura 2.4: Esempio di approccio di aggregazione e disaggregazione

Figura 2.5: Uso della correlazione per determinare la coerenza nei dati

Figura 2.6: Uso delle tecniche di smoothing

Figura 2.7: Risultati prodotti dallo smoothing

Figura 2.8: Calcolo del MAPE

Figura 2.9: Plot di un labor driver e del corrispondente errore di previsione

Figura 2.10: Confronto tra i livelli di staff determinati con i tre approcci nel primo scenario

Figura 2.11: Confronto tra i costi generati dai tre approcci nel primo scenario di riferimento

Figura 2.12: Confronto tra i livelli di staff determinati con i tre approcci nel secondo scenario

Figura 2.13: Confronto tra i costi generati dai tre approcci nel secondo scenario di riferimento

Figura 2.14: Applicazione di uno standard economico basato sulla teoria delle code

Figura 2.15: Modello della teoria delle code

Figura 2.16: Applicazione di uno standard economico

Figura 2.17: Effetti di un aumento del costo di una transazione

Figura 2.18: Identificazione di più intervalli di valutazione per stimare le vendite perdute

Figura 2.19: Determinazione del numero di impiegati da schedulare

Figura 2.20: Valutazione dei costi derivanti da uno scostamento dal numero ideale di staff

Figura 2.21: Modello di Dantzig

Figura 2.22: Modello di Keith

Figura 2. 23: Confronto tra gli approcci ad una o due fasi per un problema di scheduling

Figura 2. 24: Confronto fra i metodi per la schedulazione del lavoro controllabile

Figura 2.25: Confronto tra le vendite in quattro lunedì consecutivi

Figura 2.26: Analisi di correlazione

Figura 2. 27: Risultato dell'applicazione della tecnica di smoothing

Figura 2.28: Realizzazione nei periodi corrispondente al livello 1

Figura 2.29: Realizzazione del volume di business cumulativa

Figura 2.30: Diagramma delle realizzazioni per il livello 1 del volume di Business

Figura 2.31: Diagramma delle realizzazioni per il livello 2 del volume di business

Figura 2. 32: Procedura per la determinazione del livello di business

Figura 2.33: Probabilità che il conteggio di clienti ricada nei diversi livelli di business

Figura 2.34: Flessibilità fornita dall'utilizzo dei dipendenti cross-trained

Figura 2.35: Flessibilità fornita dall'utilizzo dei dipendenti cross-trained

Figura 3.1. Dimensionamento ottimale di una stazione di servizio

Figura 3.2. Sistema di servizio

Figura 3.3. Schema di servizio con due code e serventi in parallelo

Figura 3.4. Schema di servizio con serventi in serie

Figura 3.5. Pdf e Cdf di una variabile aleatoria esponenziale
Figura 3.6. Distribuzioni di Poisson con diversi valori λ_i
Figura 3.7. Esempio di realizzazione di arrivi Poissoniani e di Erlang-10
Figura 3.8. Rappresentazione tramite grafo orientato delle probabilità di transizione
Figura 3.9. Diagramma di transizione di stato di un processo di nascita e morte
Figura 3.10. Bilanciamento dei flussi entranti ed uscenti allo stato zero
Figura 3.11. Bilanciamento dei flussi entranti ed uscenti in un generico stato i
Figura 3.12. Esempio di sistema di servizio M/M/1
Figura 3.13. Processo di nascita e morte che rappresenta un sistema M/M/1
Figura 3.14. Periodo di Occupazione come complemento del Periodo di Inattività
Figura 3.15. Processo di nascita e morte che rappresenta un sistema M/M/s
Figura 3.16. Sistema di servizio senza alcuna forma di abbandono dei clienti
Figura 3.17. Blue print del comportamento del cliente in un sistema di servizio
Figura 3.18. Sistema di servizio con balking e reneging
Figura 3.19. Esempio di distribuzione della funzione di pazienza di un cliente
Figura 3.20. Stati d'animo di un cliente, nel tempo, in una coda visibile e una invisibile

Figura 4.1: Andamento delle vendite per i diversi segmenti di clientela.
Figura 4.2: Incidenza del controllo sulla capacità
Figura 4.3: Revenue ottenuto con una politica di discriminazione dei prezzi
Figura 4.4: Protection Level e Booking Limit
Figura 4.5: Albero decisionale per un problema di yield management
Figura 4.6: Inputs ed output del modello 1
Figura 4.7: Inputs ed output del modello 2
Figura 4.8: Andamento dell'overbooking sulle prenotazioni
Figura 4.10: Processo in due fasi
Figura 4.11: Overbooking nel tempo
Figura 4.12: Costi reali connessi all'overbooking
Figura 4.13: Andamento reale dei ricavi e dei costi
Figura 4.14: Albero decisionale per un problema di overbooking
Figura 4.15: La proprietà di Monge
Figura 4.16: Processo tipico dell'Ares
Figura 4.17: Schema di funzionamento dell'analisi sui valori medi
Figura 4.18: Schema di funzionamento dell'analisi tramite foglio di calcolo
Figura 4.19 Flow chart: fase "prenotazione"
Figura 4.20 flow chart fase: " gestione overbooking giornaliero"
Figura 4.21 flow chart fase "Annullamento prenotazione"
Figura 4.22 flow chart fase "Modifica prenotazione"
Figura 4.23 Flow chart: fase: "check in"
Figura 4.24 Flow chart: fase: "check out"
Figura 4.25 Pannello di controllo
Figura 4.26 Bottone "Ricerca disponibilità"
Figura 4.27 Bottone "Torna al pannello di controllo"
Figura 4.28 messaggio errore
Figura 4.29 messaggio errore
Figura 4.30 messaggio errore
Figura 4.31 messaggio errore
Figura 4.32 Bottone "Gestione overbooking"

Figura 4.33 Bottone “Ricerca disponibilità corsi simili”
Figura 4.34 Bottone “Modifica/ cancella prenotazioni”
Figura 4.35 Maschera “ Dati prenotazione da modificare / cancellare
Figura 4.36 messaggio informativo
Figura 4.37 Bottone “Cerca prenotazioni corrispondenti”
Figura 4.38 bottone “Torna al pannello di controllo”
Figura 4.39 Bottone “check in”
Figura 4.40 Maschera “ Check in”
Figura 4.41 Bottone “Torna al pannello di controllo”
Figura 4.42 Bottone “Apri lista check in”
Figura 4.43 Maschera “Lista check in”
Figura 4.44 messaggio informativo
Figura 4.45 Bottone “Torna a check in”
Figura 4.46 Bottone “Apri report check in”
Figura 4.47 Bottone “Identifica il cliente come no-show”
Figura 4.48 Bottone “ Check out”
Figura 4.50 Maschera “Check out”
Figura 4.51 Bottone “Torna al pannello di controllo”
Figura 4.52 Bottone “Apri elenco prenotazioni ricercate”
Figura 4.53 messaggio informativo
Figura 4.54 Maschera “Elenco prenotazioni ricercate”
Figura 4.55 messaggio informativo
Figura 4.56 Bottone “Visualizza ricevuta cliente”
Figura 4.57 Prenotazioni mediante ottimizzazione
Figura 4.58 prenotazioni in overbooking